Vol. I 1930

THE ANNALS
of
MATHEMATICAL STATISTICS

(Printed in U. S. A.)

Published and Lithoprinted by
EDWARDS BROTHERS, Inc.
ANN ARBOR, MICH

A quarterly publication sponsored by the American Statistical Association, devoted to the theory and application of Mathematical Statistics.

The rates are six dollars per annum.

Reprints of any article in this volume may be obtained at any time from the Editor at the following rates, postage included.

Number of copies	Cost per page
1- 4 . .	2 cents
5-24 . .	1½ cents
25-49 . .	1 cent
50 and over .	¾ cent

Address: Editor, Annals of Mathematical Statistics
Post Office Box 171, Ann Arbor, Michigan

THE ANNALS OF MATHEMATICAL STATISTICS

VOL. I **FEBRUARY** **NO. 1**

CONTENTS

PUBLISHED QUARTERLY BY
AMERICAN STATISTICAL ASSOCIATION

Publication Office—Edwards Brothers, Inc., Ann Arbor, Michigan
Business Office—530 Commerce Bldg., New York Univ., New York, N. Y.

Entered as second class matter at the Postoffice at Ann Arbor, Mich., under the Act of March 3rd, 1879.

A quarterly publication sponsored by the American Statistical Association, devoted to the theory and application of Mathematical Statistics.

The rates are six dollars per annum.

Reprints of any article in this issue may be obtained at any time from the Editor at the following rates, postage included.

Number of copies	Cost per page
1- 4 . .	2 cents
5-24 . .	1½ cents
25-49 . .	1 cent
50 and over .	¾ cent

ADDRESS: Editor, Annals of Mathematical Statistics
Post Office Box 171, Ann Arbor, Michigan

THE ANNALS OF MATHEMATICAL STATISTICS

By

WILLFORD I. KING

For ninety-one years the American Statistical Association has held the van in matters statistical in the United States. At the time when our Association was founded, statistical method was an extremely simple science. In recent years, the technique has, however, been growing more and more complex. The *Journal of the American Statistical Association* has served all the members of the Association and an attempt has been made to cover, in its pages, all phases of statistical method. For some time past, however, it has been evident that the membership of our organization is tending to become divided into two groups — those familiar with advanced mathematics, and those who have not devoted themselves to this field. The mathematicians are, of course, interested in articles of a type which are not intelligible to the non-mathematical readers of our Journal. The Editor of our Journal has, then, found it a puzzling problem to satisfy both classes of readers.

Now a happy solution has appeared. The Association at this time has the pleasure of presenting to its mathematically inclined members the first issue of the ANNALS OF MATHEMATICAL STATISTICS, edited by Prof. Harry C. Carver of the University of Michigan. This Journal will deal, not only with the mathematical technique of statistics, but also with the applications of such technique to the fields of astronomy, physics, psychology, biology, medicine, education, business, and economics. At present, mathematical articles along these lines are scattered through a great variety of publications. It is hoped that in the future they will be gathered together in the ANNALS.

The editorial policy will be to select articles that will best meet the needs of the time. There can be no questioning the statement that at the present time there are in this country many more who need stimulation in the fundamentals of mathematical statistics than there are individuals whose prime interest is in the advancement of modern statistical theory. Therefore particular stress will be laid on articles of a fundamental nature during the first few years of the life of the ANNALS. The officers, after due deliberation, have chosen a new method of printing in order to facilitate the composition of original articles and the obtaining of reprints. A photographic process is employed, which will permit the Association at any point in the

future to furnish reprints or back numbers. The advantages of this to libraries and classes in statistics is apparent. A particular effort will be made to insert from time to time tables that must be constantly referred to by statisticians. Nevertheless, the chronicling of research will in no sense be neglected.

My personal opinion is that the advent of the ANNALS constitutes an important milestone in the history of our Association. I am sure that this new publication will be welcomed heartily, not only by the mathematically trained section of our membership, but also by the non-mathematical group, for the latter recognize that the more advanced phases of mathematics are rendering extremely valuable service in furthering the progress of statistical technique, thus aiding in the solution of problems of the greatest moment.

REMARKS ON REGRESSION

By

S. D. Wicksell

1. In a paper published twelve years ago[1] I derived a set of formulae for bivariate regression which were found to give good results on unimodal materials of a fairly general nature and which in the case of moderately skew distributions, were reduced to very simple and easily applicable forms. Two years later I extended the theory also to the case of multiple correlations of similar types[2]. These formulae were deduced on the assumption that the correlation surface could be expressed by a so-called series of type A[3], i. e. that the deviations from the best fitting normal surface could be expressed as a series, developed according to the derivatives of different orders of the Bravais function, expressing that normal surface.

When, after the lapse of so many years, I find that this theory has not received the attention which it seems to me it merits in view of the very simple, and on a fairly large class of curved regressions readily applicable results, I attribute this in part at least to the apparent (not actual) speciality of the assumptions made with regard to the mathematical expression for the correlation surface, and in part also to the rather repellent show of mathematics involved in the deductions. In the hope to give the theory a better chance of coming to the attention of statisticians, I propose here to deduce some of my main results in an entirely different way, bringing the theory back on more simple principles. I believe that by this method of deduction it will be more easy for the reader to see exactly where assumptions come in, and also the nature of the restrictions caused by these assumptions.

2. Let x and y be a pair of correlated variates, our material

---

1. The correlation function of Type A, and the regression of its characteristics. Kungl. Svenska Vetenskapsakademiens Handlinger Bd. 58 Nr 3 1917 Also "Meddelanden fran Lunds Astronomiska Observatorium" Ser II Nr 17
2. Multiple correlation and non-linear regression. Arkiv. for Matematik, Fysik och Astronomi. Bd 14 Nr. 10, 1919. Also "Meddelanden fran Lunds Astronomiska Observatorium." Ser. I, Nr. 91.
3. Charlier. Contributions to the mathematical theory of statistics. 6. The correlation function of type A. Arkiv for Matematik, Fysik och Astronomi. Bd. 9, Nr. 26, 1914. Also "Meddelanden fran Lunds Astronomiska Observatorium" Ser. I. Nr. 58.

consisting of N such pairs. Computing the means and central moments, we have

$$M_x = \frac{1}{N}\sum x \; ; \quad M_y = \frac{1}{N}\sum y \; ; \quad \mu_{ij} = \frac{1}{N}\sum (x-M_x)^i (y-M_y)^j$$

The standard deviations of x and y and the coefficient of correlation are then defined by

$$\sigma_x = \sqrt{\mu_{20}} \quad ; \qquad \sigma_y = \sqrt{\mu_{02}} \quad ; \qquad r = \frac{\mu_{11}}{\sigma_x \sigma_y}$$

Following Yule[1] and Pearson[2] we now treat the problem of regression as a simple problem of graduation, defining the regression of y on x as a parabola of a given degree, which, with x as argument, is fitted to the y's by the method of least squares. The regression may then be written in the form

$$y_x - M_y = a_0 + a_1 (x - M_x) + a_2 (x - M_x)^2 + \cdots + a_p (x - M_x)^p,$$

and the least squares normal equations for determining the parameters $a_0, a_1, a_2, \cdots a_p$ assume the form (Pearson Op. Cit. p. 25).

$$(1)\quad \begin{cases} 0 = a_0 + a_2\mu_{20} + a_3\mu_{30} + \cdots + a_p\mu_{p,0} \\ \mu_{11} = a_1\mu_{20} + a_2\mu_{30} + a_3\mu_{40} + \cdots + a_p\mu_{p+1,0} \\ \mu_{21} = a_0\mu_{20} + a_1\mu_{30} + a_2\mu_{40} + a_3\mu_{50} + \cdots + a_p\mu_{p+2,0} \\ \mu_{31} = a_0\mu_{30} + a_1\mu_{40} + a_2\mu_{50} + a_3\mu_{60} + \cdots + a_p\mu_{p+3,0} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \mu_{p,1} = a_0\mu_{p,0} + a_1\mu_{p+1,0} + a_2\mu_{p+2,0} + a_3\mu_{p+3,0} + \cdots + a_p\mu_{2p,0} \end{cases}$$

---

1. On the Theory of Correlation. Jour. Roy. Stat. Soc., Vol. 60, 1897, and On the Theory of Correlation for any number of Variables treated by a new System of Notation. Proc. Roy. Soc., Ser. A, Vol. 79, 1907.
2. Mathematical Contributions to the Theory of Evolution XIV. On the General Theory of Skew Correlation and non-Linear Regression. Drapers Co. Research Memoirs Biometric Series II. Cambridge Univ. Press, 1905.

Writing the solution in the form of determinants, we have

$$a_{i-1} = \frac{1}{\Delta} \cdot \Delta_i ,$$

where

$$(3) \qquad \Delta = \begin{vmatrix} 1, & 0, & \mu_{20}, & \mu_{30}, & \cdots & \mu_{p,0} \\ 0, & \mu_{20}, & \mu_{30}, & \mu_{40}, & \cdots & \mu_{p+1,0} \\ \mu_{20}, & \mu_{30}, & \mu_{40}, & \mu_{50}, & \cdots & \mu_{p+2,0} \\ \mu_{30}, & \mu_{40}, & \mu_{50}, & \mu_{60}, & \cdots & \mu_{p+3,0} \\ \cdot & \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ \mu_{p,0}, & \mu_{p+1,0}, & \mu_{p+2,0}, & \mu_{p+3,0}, & \cdots & \mu_{2p,0} \end{vmatrix}$$

and Δ_i is obtained when the i'th row in Δ is exchanged for the left membra of equations (1), i. e. for the series of elements:

$0, \ \mu_{11}, \ \mu_{21}, \ \mu_{31}, \ \cdots \ \mu_{p,1}.$

3. Some important general conclusions may at once be derived from this system. Defining as *non-regression of the p'th order* the case that all the coefficients $a_1, a_2, a_3, \cdots a_p$ turn out to be practically equal to zero, i. e. that a horizontal straight line is the best parabola of the p'th degree that can be fitted to the series of y's, it is first seen, from the first of equations (1), that then also $a_0 = 0$. Secondly we can draw the conclusion that this can take place only if all the elements $\mu_{11}, \ \mu_{21}, \ \mu_{31}, \ \cdots \ \mu_{p,1}$, are equal to zero. Hence the condition for non-regression of the p'th order of y on x is that we have

$$(4) \qquad \mu_{i,0} = 0 \quad for \ i = 1, 2, 3, \cdots p$$

This clearly involves also that the coefficient of correlation, r equals zero.

Defining further as *linear regression of the p'th order* the case that the coefficients $a_2, a_3, \cdots a_p$ are equal to zero, i. e. that a non-horizontal straight line is the best parabola of the p'th degree that can be fitted to the series of y's, we immediately see, from the two first of equations (1), that then we must have

$$(5) \qquad a_0 = 0 ; \qquad a_1 = \frac{\mu_{11}}{\mu_{20}}$$

Referring here to the well-known theorem that any determinant will disappear, when the elements of two rows are proportional (the elements of any one row being obtained by multiplying the corresponding elements of another row by a constant factor) it is easily seen that all the determinants Δ_i except Δ_2, and hence by (2) all the coefficients $a_0, \cdots a_p$, except a_1, will disappear if the quantities $0, \mu_{11}, \mu_{21}, \mu_{31}, \cdots \mu_{p,1}$, in the left membra of (1) are proportional to the elements $0, \mu_{20}, \mu_{30}, \cdots \mu_{p+1,0}$ in the second row of the determinant Δ. Hence the condition for linear regression of the p'th order of y on x is that we have

$$(6) \qquad \mu_{11} \cdot \mu_{i+1,0} = \mu_{20}\, \mu_{i,1} \qquad for\ i = 1, 2, 3, \cdots p.$$

A few considerations will show that this condition is not only sufficient but also necessary. For $p = 3$ these criteria were demonstrated by Pearson.

4. Thus far there are no other assumptions involved than the principle of least squares, and that the regression of y on x may be described by a whole rational function. The chief difficulty in the application of this theory of regression is that, as seen from equation (1), in order to determine a regression of the p'th degree we must compute and use moments (of the series of x's up to the order $2p$). Now, as justly remarked by Pearson, moments of high orders are, on account of their large standard errors, very little to be relied upon, at least in the case of ordinary materials (N not very large) . Besides this, the numerical labor involved in computing higher moments is comparatively very great. Hence, Pearson's theory of regression will be practically applicable only in cases when the regression is at the most parabolic of the second degree. Indeed, this is a very serious restriction, because curved regressions often have at least one inflection. Thus in order to meet fairly frequent cases of regression we must needs have recourse at least to cubic parabolas. But this should require the computation of all the moments of x up to the sixth order.

In order to remove, as far as possible, this difficulty, I take refuge in a golden rule expressed by Thiele[1]. Thiele introduces, instead of the moments, a system of coefficients called the semi-invariants. These semi-invariants (here denoted by $\lambda_{i,0}$) are defined in terms of the moments by the identity:

1. Theory of Observations. London 1903, p. 49.

$$\lambda_{20}\frac{x^2}{2!}+\lambda_{30}\frac{x^3}{3!}+\lambda_{40}\frac{x^4}{4!}+\cdots$$
$$=\log_e\left(1+\mu_{20}\frac{x^2}{2!}+\mu_{30}\frac{x^3}{3!}+\mu_{40}\frac{x^4}{4!}+\cdots\right.$$

Developing, we find

$$(7)\quad \begin{aligned}&\lambda_{20}=\mu_{20};\ \lambda_{30}=\mu_{30};\ \lambda_{40}=\mu_{40}-3\mu_{20}^2:\\&\lambda_{50}=\mu_{50}-10\mu_{30}\mu_{20};\lambda_{60}=\mu_{60}-15\mu_{40}\mu_{20}+30\mu_{20}^3-10\mu_{30}^2\end{aligned}$$

Now, the rule indicated by Thiele is the following:

To obtain the first semi-invariants rely entirely on computations. To obtain the intermediate semi-invariants rely partly on computations, partly on theoretical considerations. But to obtain the higher semi-invariants rely entirely on theoretical considerations.

Of course, this rule is just as well applicable to the determination of moments, as any moment may be expressed in terms of the semi-invariants of the same and lower order. In particular we have

$$(8)\quad \begin{aligned}&\mu_{20}=\lambda_{20};\ \mu_{30}=\lambda_{30};\ \mu_{40}=\lambda_{40}+3\lambda_{20}^2:\\&\mu_{50}=\lambda_{50}-10\lambda_{30}\lambda_{20};\mu_{60}=\lambda_{60}+15\lambda_{40}\lambda_{20}+15\lambda_{20}^3+10\lambda_{30}^2\end{aligned}$$

5. A most natural way of applying the rule is afforded by Pearson's celebrated theory of frequency-functions. The moments $\mu_{i,0}$ are the moments of one of the marginal distributions (here the distribution of the x's). Computing μ_{20}, μ_{30} and μ_{40} in the ordinary way from the observations, criteria can be formed[1] showing to which of the Pearson Types the frequency curve of x belongs. This being decided, the parameters of the curve may be determined by the aid of the same moments. As the moments of higher order are easily expressed in terms of the parameters we get, in this way, μ_{50} and μ_{60} expressed in terms of μ_{20}, μ_{30} and μ_{40}.

To state the matter in a more general way, we may use the formulae given by Pearson in his memoir on regression, loc. cit. pp. 5 and 6.

1. See W. Palin Elderton: Frequency Curves and Correlation. London 1927, Table VI.

Pearson starts from a differential equation of the form

$$(9)\quad f'(x)(b_0+b_1x+b_2x^2+b_3'x^3+\cdots)=(x+a)f(x)$$

where $f(x)$ is the frequency function of x.

Multiplying on both sides by x and integrating by parts, he finds the following formulae[1] (placing the origin in the mean)

$$(10)\quad nb_0\mu_{n-1,0}+(n+1)b_1\mu_{n,0}+(n+2)\,b_2\,\mu_{n+1,0}+\cdots$$
$$=-\mu_{n+1,0}-a\mu_{n,0}$$

Now, Pearson remarks that experience shows that for the great bulk of frequency distributions the higher terms, multiplied by b_3, b_4, etc., may be neglected. In fact, Pearson's system of frequency curves is obtained as a result of putting $b_i=0$ for $i\geq 3$.

Following Pearson's example, we get the recursion formula,

$$(11)\quad nb_0\mu_{n-1,0}+\left[(n+1)\,b_1+a\right]\mu_{n,0}=-\left[(n-2)\,b_2-1\right]\mu_{n+1,0}$$

Putting here $n=0, 1, 2, 3$, we get four equations to determine a, b_0, b_1, and b_2 in terms of the moments μ_{20}, μ_{30}, and μ_{40}. This being done, we get μ_{50} and μ_{60} on putting $n=4$ and 5.

The procedure indicated above leads, in fact, to the theory of skew regression which is the natural consequence of Pearson's theory of skew frequency curves.

6. As the theory just indicated above is at present at my request being worked out in detail by one of my pupils, Mr. Walter Anderson, I refrain from proceeding further into the matter.

It remains, however, to show how the special formulae for cubic regression, given by me twelve years ago, arise out of a somewhat similar procedure.

Instead of starting from Pearson's theory of frequency functions, I now start from Thiele's theory of frequency functions. Just as in the preceding section the coefficients b_3, b_4 etc. were neglected in the equation (10), given by Pearson, I now neglect the semi-invariants λ_{50} and λ_{60} in the equations (8), given by Thiele. There is no doubt that the former approximation is of

1. See also Palin Elderton, Op. cit. p. 39.

far more general validity than the latter; still the latter may be justified by the following considerations.

Assuming the variate x to be generated as the sum of a large number of independent, elementary increments, each of which has its own frequency distribution and its own set of semi- invariants, it follows from the theory of Thiele that any semi-invariant $\lambda_{r.o}$ of x is the sum of the elementary semi-invariants of the same order. Supposing the elementary increments to be s in number and denoting by λ'_r the mean value of the r elementary semi-invariants of order r we consequently have

$$\lambda_{r,} = s\lambda'_r$$

Hence we get

$$\gamma_{r,o} = \frac{\lambda_{r,o}}{\lambda_{20}^{r/2}} = \frac{\lambda'_r}{\lambda_2^{r/2}} \; \frac{1}{s^{\frac{r-2}{2}}}$$

Except under rather special conditions, which it is not necessary to dwell on here, the ratios $\lambda'_r/\lambda'^{r/2}_2$ are not extensively great. Thus if s is a large number we see that the "standardized" semi-invariants $\gamma_{r,o}$ of x are small of the order of magnitude of $\left(\frac{1}{\sqrt{s}}\right)^{r-2}$ In particular we have.

γ_{30}	of the	order		$\frac{1}{\sqrt{s}}$
γ_{40}	"	"	"	$\frac{1}{s}$
γ_{50}	"	"	"	$\frac{1}{s\sqrt{s}}$
γ_{60}	"	"	"	$\frac{1}{s^2}$

We now have, denoting by

$$\alpha_{r,o} = \frac{\mu_{r,o}}{\mu_{20}^{r/2}}$$

the "standardized" moment of x, by a simple transformation of equation (8).

$$(8') \quad \alpha_{20} = 1\,; \quad \alpha_{30} = \gamma_{30}\,; \quad \alpha_{40} = \gamma_{40} + 3\,;$$
$$\alpha_{50} = \gamma_{50} + 10\gamma_{30}\,; \quad \alpha_{60} = \gamma_{60} + 15\gamma_{40} + 10\gamma_{30}^2 + 15$$

Stopping with quantities of the order $\frac{1}{s}$ we get

$$(13) \quad \alpha_{50} = 10\gamma_{30}\,; \quad \alpha_{60} = 15\gamma_{40} + 10\gamma_{30}^2 + 15$$

In practice we can, of course, not very well know if the hypothesis of elementary increments is valid, but if we have, on computing the moments up to the fourth order, found that γ_{30} and γ_{40} are rather small, and that γ_{40} is of the order of magnitude of γ_{30}^2, there is a certain plausibility in assuming that γ_{50} and γ_{60} are still smaller and that they may be neglected as compared to γ_{40} and γ_{30}^2.

The curve of cubic regression of y on x we may write in the form

$$t_y = c_0 + c_1 t_x + c_2 t_x^2 + c_3 t_x^3$$

where we have put

$$t_x = \frac{x - M_x}{\sqrt{\mu_{20}}} \quad ; \qquad t_y = \frac{y_x - M_y}{\sqrt{\mu_{20}}}$$

and it is evident that equation (1) now takes the form

$$\begin{aligned}
0 &= c_0 &&+ c_2 &&+ c_3\alpha_{30} \\
r &= &+ c_1 &+ c_2\alpha_{30} &&+ c_3\alpha_{40} \\
\alpha_{21} &= c_0 &+ c_1\alpha_{30} &+ c_2\alpha_{40} &&+ c_3\alpha_{50} \\
\alpha_{31} &= c_0\alpha_{30} &+ c_1\alpha_{40} &+ c_2\alpha_{50} &&+ c_3\alpha_{60}
\end{aligned}$$

We get

$$\begin{aligned}
(14)\ \Delta = {} & \alpha_{60}(\alpha_{40} - \alpha_{30}^2 - 1) - \alpha_{50}(\alpha_{50} - 2\alpha_{30}\alpha_{40} - 2\alpha_{30}) \\
& - \alpha_{40}(\alpha_{40}^2 - \alpha_{40} + 3\alpha_{30}^2) + \alpha_{30}^4
\end{aligned}$$

$$\begin{aligned}
\Delta_1 = {} & r(\alpha_{30}\alpha_{60} - \alpha_{50}\alpha_{40}) - \alpha_{21}(\alpha_{60} - \alpha_{40}^2) + \alpha_{31}(\alpha_{50} - \alpha_{40}\alpha_{30}) \\
& - r\alpha_{30}(\alpha_{30}\alpha_{50} - \alpha_{40}^2) + \alpha_{21}\alpha_{30}(\alpha_{50} - \alpha_{30}\alpha_{40}) - \alpha_{31}\alpha_{30}(\alpha_{40} - \alpha_{30}^2)
\end{aligned}$$

$$\begin{aligned}
\Delta_2 = {} & r(\alpha_{40}\alpha_{60} - \alpha_{50}^2 - \alpha_{60} + 2\alpha_{30}\alpha_{50} - \alpha_{30}^2\alpha_{40}) \\
& - \alpha_{21}(\alpha_{30}\alpha_{60} - \alpha_{40}\alpha_{50} + \alpha_{30}\alpha_{40} - \alpha_{30}^3) + \alpha_{31}(\alpha_{30}\alpha_{50} - \alpha_{40}^2 + \alpha_{40} + \alpha_{30}^2)
\end{aligned}$$

$$\begin{aligned}
\Delta_3 = {} & -r(\alpha_{30}\alpha_{60} - \alpha_{50}\alpha_{40} + \alpha_{40}\alpha_{30} - \alpha_{30}^2) + \alpha_{21}(\alpha_{60} - \alpha_{40}^2 - \alpha_{30}^2) \\
& - \alpha_{31}(\alpha_{50} - \alpha_{30}\alpha_{40} - \alpha_{30})
\end{aligned}$$

$$\Delta_4 = r(\alpha_{50}\alpha_{30} - \alpha_{40}^2 + \alpha_{40} - \alpha_{30}^2) - \alpha_{21}(\alpha_{50} - \alpha_{30}\alpha_{40} - \alpha_{30}) + \alpha_{31}(\alpha_{40} - \alpha_{30}^2 - 1)$$

And the coefficients are

$$c_0 = \frac{\Delta_1}{\Delta} \qquad c_1 = \frac{\Delta_2}{\Delta} \qquad c_2 = \frac{\Delta_3}{\Delta} \qquad c_4 = \frac{\Delta_4}{\Delta}$$

We now introduce the semi-invariants by (8'), taking for α_{50} and α_{60} the approximate formulae (13). For α_{21} and α_{31} we put

$$(15) \qquad \alpha_{21} = \gamma_{21} \quad ; \quad \alpha_{31} \doteq \gamma_{31} + 3r$$

The coefficients γ_{21} and γ_{31} are then the standardized correlation semi-invariants, according to a generalized theory of semi-invariants for bi-variate distributions.

It is now a consequence of our principle ot approximation that all powers and products $\gamma_{ij}, \gamma_{k,l}, \gamma_{m,n} \ldots$, of which the sum $i+j+k+l+m+n+\ldots$ of the indices exceeds 6, shall be neglected as compared to powers and products of lower order. Observing this, the determinants reduce to the following:

$$\Delta = 12(1 - 2\gamma_{30}^2 + 2\gamma_{40}),$$

or

$$\frac{1}{\Delta} - \frac{1}{72}(1 + 2\gamma_{30}^2 - 2\gamma_{40}),$$

$$\Delta_1 = 6(r\gamma_{30} - \gamma_{21}),$$

$$\Delta_2 = 12r + 6(r\gamma_{40} - \gamma_{31}) + 24r\gamma_{40} - 24r\gamma_{30}^2 - 12\gamma_{30}(r\gamma_{30} - \gamma_{21}),$$

$$\Delta_3 = -6(r\gamma_{30} - \gamma_{21}),$$

$$\Delta_4 = -2(r\gamma_{40} - \gamma_{31}) + 6\gamma_{30}(r\gamma_{30} - \gamma_{21}).$$

Using the same rule of approximation on multiplying by $\frac{1}{\Delta}$, we finally get

$$
(16)\qquad
\begin{aligned}
c_0 &= \frac{1}{2}(r\gamma_{30} - \gamma_{21}),\\
c_1 &= r + \frac{1}{2}(r\gamma_{40} - \gamma_{31}) - \gamma_{30}(r\gamma_{30} - \gamma_{21}),\\
c_2 &= -\frac{1}{2}(r\gamma_{30} - \gamma_{21}),\\
c_3 &= -\frac{1}{6}(r\gamma_{40} - \gamma_{31}) + \frac{1}{2}\gamma_{30}(r\gamma_{30} - \gamma_{21}).
\end{aligned}
$$

In my cited memoir of twelve years ago I put[1]

$$r_{30} = \frac{1}{2}(r\gamma_{30} - \gamma_{21}); \qquad r_{40} = -\frac{1}{6}(r\gamma_{40} - \gamma_{31}).$$

Using this notation, we get

$$
(17)\qquad
\begin{aligned}
c_0 &= r_{30},\\
c_1 &= r - 3r_{40} - 2\gamma_{30}\, r_{30},\\
c_2 &= -c_0 = -r_{30},\\
c_3 &= r_{40} + \gamma_{30}\, r_{30}.
\end{aligned}
$$

These coefficients are exactly the same as in equation (34*, II) of my former memoir. As shown in that memoir on several numerical examples, the regression formula in question applies very well in cases of moderately skew correlations.

It is seen that the coefficients r_{30} and r_{40} determine the curvature of the regression. If $r_{30} = r_{40} = 0$ the regression is linear (of the third order). I have called these coefficients the correlation coefficients of higher order. If the correlation surface is approximately normal we have the following formulae for the standard errors of the coefficients involved:

1. In Pearson's notation we have $r_{30} = \frac{1}{2}\epsilon$ and $r_{40} = \frac{1}{6}\zeta$.

$$\sigma_{(\gamma_{30})} = \sqrt{\frac{6}{N}}\,; \quad \sigma_{(\gamma_{40})} = \sqrt{\frac{24}{N}}\,; \quad \sigma_{(\gamma_{21})} = \sqrt{\frac{2+4r^2}{N}}\,; \quad \sigma_{(\gamma_{31})} = \sqrt{\frac{6+18r^2}{N}}$$

$$(18) \quad \sigma_{(r)} = \frac{1-r^2}{\sqrt{N}}\,; \quad \sigma_{(r_{30})} = \sqrt{\frac{1-r^2}{2N}}\,; \quad \sigma_{(r_{40})} = \sqrt{\frac{1-r^2}{6N}}\,; \quad \sigma_{(c_0)} = \sqrt{\frac{1-r^2}{2N}}\,;$$

$$\sigma_{(c_1)} = \sqrt{\frac{(5-2r^2)(1-r^2)}{2N}}\,; \quad \sigma_{(c_2)} = \sqrt{\frac{1-r^2}{2N}}\,; \quad \sigma_{(c_3)} = \sqrt{\frac{1-r^2}{6N}}$$

Lund (Sweden).

SYNOPSIS OF ELEMENTARY MATHEMATICAL STATISTICS*

By

B. L. Shook

Section I. Elementary Statistical Functions

1. *Variates.* Practically all statistical data[1] is obtained as the result of observations that endeavor to establish the magnitudes of certain variables. The individual magnitudes that are recorded are termed variates. Thus in computing the average annual rainfall of a region, the variable is rainfall, and the amount of rainfall for any single year is a variate. Likewise, if the bank clearings for the City of New York be under consideration, then the variable is bank clearings, and the clearings for any specified interval is a variate.

2. The *arithmetic mean* of a series of variates is equal to the sum of the variates divided by the number of variates in the series. If M_v designates the arithmetic mean of the N variates v_1, v_2, v_3, . . . , v_N,

(1) $$M_v = \frac{1}{N}(v_1 + v_2 + \cdots + v_N) = \frac{1}{N}\sum v$$

3. The n th moment of a series of variates is defined as the arithmetic mean of the n th powers of these variates and is represented by the symbol $\mu'_{n:v}$. Thus ,

(2) $$\mu'_{n:v} = \frac{1}{N}(v_1^n + v_2^n + v_3^n + \cdots + v_N^n) = \frac{1}{N}\sum v^n$$

That is

$$\mu'_{1:v} = \frac{1}{N}\sum v$$
$$\mu'_{2:v} = \frac{1}{N}\sum v^2$$
$$\mu'_{3:v} = \frac{1}{N}\sum v^3$$

---

* An abstract of a series of lectures on elementary statistics given by the mathematical statistical staff at the University of Michigan.

1. Observe that the number of variates in a series is denoted by N, whereas the smaller italic n is employed as an ordinal number.

Obviously, by definitions (1) and (2)

(3) $$\mu'_{1:v} = M_v$$

4. *The deviation of a variate from the arithmetic mean* will be designated by the symbol $\bar{v}$, i. e.

(4) $$\bar{v}_i = v_i - M_v$$

5. The *nth moment about the mean** is defined as the arithmetic mean of the nth powers of the deviations of the variates from the mean, and is represented symbolically by $\mu_{n:v}$. Thus

(5) $$\mu_{n:v} = \frac{1}{N}\sum \bar{v}^n$$ so that

(5a) $$\mu_{1:v} = \frac{1}{N}\sum \bar{v} = 0$$

(5b) $$\mu_{2:v} = \frac{1}{N}\sum \bar{v}^2$$

(5c) $$\mu_{3:v} = \frac{1}{N}\sum \bar{v}^3$$

The fact that $\mu_{1:v} = 0$, is demonstrated as follows:

$$\begin{array}{rcl} \bar{v}_1 & = & v_1 - M_v \\ \bar{v}_2 & = & v_2 - M_v \\ \cdot & \cdot & \cdot \\ & \cdot & \\ \bar{v}_N & = & v_N - M_v \\ \hline \sum \bar{v} & = & \sum v - NM_v \end{array}$$

$$\mu_{1:v} = \frac{\sum \bar{v}}{N} = \frac{\sum v}{N} - M_v = M_v - M_v = 0$$ Q. E. D.

The numerical example of Table I illustrates the definitions of the preceding paragraphs. The data consists of thirteen variates, which represent the number of even numbers found in consecutive blocks of 100 numbers, drawn to determine the order of call for draft-

* For convenience the arithmetic mean is frequently referred to as *the* mean. When referring to geometric or harmonic means, the adjectives geometric or harmonic must therefore be specified.

ing United States soldiers in 1918. These variates were obtained from the *first* 1300 drawings made.

The most obvious conclusion to be drawn from Table I is that the use of fractions in determining the values of $\mu_{n:v}$ is cumbersome. if M_v is a whole number, then the values of $\bar{v}$, $\bar{v}^2$ and $\bar{v}^3$ are integers, and the procedure is simple. Generally, however, M_v will be fractional, and consequently awkward expressions for $\bar{v}$, $\bar{v}^2$ and $\bar{v}^3$ will result. On the other hand, the computation of values of $\mu'_{n:v}$ is relatively easy, and hence it is expedient to express $\mu_{2:v}$ and $\mu_{3:v}$ in terms of the moments $\mu'_{n:v}$. This may be done as follows:

Since by definition,

$$\bar{v}_i = v_i - M_v, \quad \text{it follows that}$$

$$\bar{v}_i^2 = v_i^2 - 2v_i M_v + M_v^2, \quad \text{and}$$

$$\bar{v}_i^3 = v_i^3 - 3v_i^2 M_v + 3v_i M_v^2 - M_v^3$$

Consequently

$$\begin{array}{ll}
\bar{v}_1^2 = v_1^2 - 2M_v v_1 + M_v^2 & \bar{v}_1^3 = v_1^3 - 3v_1^2 M_v + 3v_1 M_v^2 - M_v^3 \\
\bar{v}_2^2 = v_2^2 - 2M_v v_2 + M_v^2 & \bar{v}_2^3 = v_2^3 - 3v_2^2 M_v + 3v_2 M_v^2 - M_v^3 \\
\bar{v}_3^2 = v_3^2 - 2M_v v_3 + M_v^2 & \bar{v}_3^3 = v_3^3 - 3v_3^2 M_v + 3v_3 M_v^2 - M_v^3 \\
\vdots & \vdots \\
\bar{v}_N^2 = v_N^2 - 2M_v v_N + M_v^2 & \bar{v}_N^3 = v_N^3 - 3v_N^2 M_v + 3v_N M_v^2 - M_v^3 \\
\hline
\sum \bar{v}^2 = \sum v^2 - 2M_v \sum v + NM_v^2 & \sum \bar{v}^3 = \sum v^3 - 3M_v \sum v^2 + 3M_v^2 \sum v - NM_v^3
\end{array}$$

Dividing both sides of these equations through by N yields, respectively

$$\frac{\sum \bar{v}^2}{N} = \frac{\sum v^2}{N} - 2M_v \cdot M_v + M_v^2$$
$$\frac{\sum \bar{v}^3}{N} = \frac{\sum v^3}{N} - 3M_v \frac{\sum v^2}{N} + 3M_v^2 \cdot M_v - M_v^3$$

Hence

$$(6) \qquad \begin{cases} \mu_{2:v} = \mu'_{2:v} - M_v^2 \\ \mu_{3:v} = \mu'_{3:v} - 3M_v \mu'_{2:v} + 2M_v^3 \end{cases}$$

TABLE I

i	v_i	v_i^2	v_i^3	$\bar{v}_i = v_i - M_v$	$\bar{v}_i^2$	$\bar{v}_i^3$
1	51	2601	132651	− 3/13	9/169	− 27/2197
2	49	2401	117649	− 29/13	841/169	− 24389/2197
3	60	3600	216000	114/13	12996/169	1481544/2197
4	53	2809	148877	23/13	529/169	12167/2197
5	48	2304	110592	− 42/13	1764/169	− 74088/2197
6	51	2601	132651	− 3/13	9/169	− 27/2197
7	42	1764	74088	− 120/13	14400/169	− 1728000/2197
8	50	2500	125000	− 16/13	256/169	− 4096/2197
9	51	2601	132651	− 3/13	9/169	− 27/2197
10	52	2704	140608	10/13	100/169	1000/2197
11	54	2916	157464	36/13	1296/169	46656/2197
12	53	2809	148877	23/13	529/169	12167/2197
13	52	2704	140608	10/13	100/169	1000/2197
Total	666	34314	1777716	0	32838/169	− 276120/2197

$$M_v = \mu'_{1:v} = \frac{666}{13} \qquad \mu_{1:v} = 0$$

$$\mu'_{2:v} = \frac{34314}{13} \qquad \mu_{2:v} = \frac{32838}{13 \cdot 169} = \frac{2526}{169}$$

$$\mu'_{3:v} = \frac{177716}{13} \qquad \mu_{3:v} = \frac{-276120}{13 \cdot 2197} = \frac{-21240}{2197}$$

These formulae are perhaps the most important in our work, since they enable us to obtain the moments about the mean without requiring that we actually determine the deviations. Applying these formulae to the numerical example of Table I,

$$\mu_{2:v} = \frac{34314}{13} - \left(\frac{666}{13}\right)^2 = \frac{2526}{169}$$

$$\mu_{3:v} = \frac{177716}{13} - 3\left(\frac{34314}{13}\right)\left(\frac{666}{13}\right) + 2\left(\frac{666}{13}\right)^3 = -\frac{21240}{2197}$$

The results thus obtained by this *indirect* method are identical with the results obtained in Table I by employing the *direct* method.

7. *Standard Deviation.* The second moment about the mean, $\mu_{2:v}$, is a function of the variability of the data, since its essential elements are the deviations of the variates from the mean. But if the original variates happen to be measured in *inches*, then since $\mu_{2:v}$ is the average of the squares of the deviations, it follows that the unit of $\mu_{2:v}$ is square inch. Nevertheless, by extracting the square root of $\mu_{2:v}$ we would obtain a function which would in general measure the variability of, and possess the same unit as the original data. This function is known as the *standard deviation* and is denoted by the symbol σ_v. Thus

$$\sigma_v = \sqrt{\mu_{2:v}} \tag{7}$$

Verbally we may say that the standard deviation is defined as the square root of the mean of the squared deviations of the variates from their mean.

Actually σ_v is rarely computed directly from the squared deviations, but rather by employing the relationship given in formula (6). For the data of Table I

$$\sigma_v = \sqrt{\frac{2526}{169}} = \frac{50.2593}{13} = 3.78918$$

8. *Standard Units.* If we assume that the arithmetic mean and the standard deviation of the weights of adult males are 150 lbs. and 20 lbs. respectively, then we may say that a man weighing 190 lbs. is

40 lbs. or 2 *standard units* above the average in weight. Likewise an individual weighing 120 lbs. may be considered as being 30 lbs. or 1.5 *standard units* under average weight. Conversely, if the arithmetic mean and the standard deviation for heights be 67 inches and 2.5 inches respectively, then an individual who is 2 standard units above the average height must be five inches above the average stature, or in other words must be 72 inches tall. The magnitude of an observation expressed in standard units is therefore defined as follows:

$$t_i = \frac{v_i - M_v}{\sigma_v} = \frac{\bar{v}_i}{\sigma_v} \tag{8}$$

It will be observed that these *standard* variates, t_i, are abstract numbers. For example, if the original variates be expressed in the unit inch then the unit of M_v, $\bar{v}$ and σ_v is also inch, and it follows that if both the numerator and denominator of a fraction be expressed as *inches* the quotient must be an abstract number, *independent of the unit employed in the measurements.* For instance, one series of variates would result if the height of each of a group of individuals were recorded in *inches*. However, if their heights had been recorded in *centimeters*, each of the resulting set of variates would be numerically about 2.54 times as large as the corresponding variate expressed in inches. Nevertheless, the *standard variates* obtained by both methods would agree in the case of each individual. Thus, if

$$M_v = 67 \text{ ins.} = 67(2.54) \text{ cms.},$$

and

$$\sigma_v = 2.5 \text{ ins.} = 2.5(2.54) \text{ cms.},$$

then for an individual 6 feet tall

$$v = 72 \text{ ins.} = 72(2.54) \text{ cms.},$$

$$\bar{v} = 5 \text{ ins.} = 5(2.54) \text{ cms.},$$

$$t = \frac{5 \text{ ins.}}{2.5 \text{ ins.}} = \frac{5(2.54) \text{ cms.}}{2.5(2.54) \text{ cms.}}, \text{ or}$$

$$t = 2 = 2.$$

With the aid of a computing machine, the series of standard variates corresponding to any observed series of variates may be completed very rapidly by means of a so-called continuous process. To

illustrate, we found that for the data of Table I, page 17,

$$M_v = 51.230769$$

$$\sigma_v = 3.86610$$

By formula (8), then

$$t_i = \frac{v_i - 51.230769}{3.86610} = -13.2513 + .258659\ v_i$$

In using this equation one should first subtract out 13.2513 from the machine, and then set up .258659 as a multiplier. The product of this multiplier by 51 will cause the value $t = -.059691$ to appear on the machine. By merely subtracting the multiplier two times, the value $t = -.577009$, corresponding to $t = 49$, appears. Continuing this "build-over" method, the following set of standard variates is readily obtained:

TABLE II

i	v_i	t_i
1	51	- .06
2	49	- .58
3	60	2.27
4	53	.46
5	48	- .84
6	51	- .06
7	42	- 2.39
8	50	- .32
9	51	- .06
10	52	.20
11	54	.72
12	53	.46
13	52	.20
Total	666	0.00

It is scarcely an exaggeration to state that the theory of mathematical statistics hinges on standard units. Although in many problems this might not appear on the surface, yet we shall see that the fact is nevertheless true.

9. The properties of the *moments* of standard variates are ....

interesting and important. Thus

(9) $$\mu_{1:t} = M_t = 0$$

since

$$M_t = \frac{\sum t}{N} = \frac{1}{N}\sum \frac{v_i - M_v}{\sigma_v} = \frac{1}{N\sigma_v}\sum \bar{v}_i = 0 .$$

(see formula 5a)

Referring to formula (6) we see that

$$\mu_2 = \mu_2' - M^2$$

$$\mu_3 = \mu_3' - 3\mu_2' M + 2M^3$$

But since M_t has already been proven equal to 0,

$$\mu_{2:t}' = \mu_{2:t}$$

$$\mu_{3:t}' = \mu_{3:t}$$

Which is an important simplification in the moments of the standard variates.

(10) $$\mu_{2:t} = 1$$

for $$\mu_{2:t} = \frac{\sum t^2}{N} = \frac{1}{N}\sum\left(\frac{v_i - M_v}{\sigma_v}\right)^2 = \frac{1}{\sigma_v^2}\sum\frac{\bar{v}^2}{N}$$

$$= \frac{\mu_{2:v}}{\sigma_v^2} = 1$$ (see formula 7)

(11) $$\mu_{3:t} = \frac{\mu_{3:v}}{\sigma_v^3} = \frac{\mu_{3:v}}{\mu_{2:v}\sigma_v}$$

for

$$\mu_{3:t} = \frac{\sum t^3}{N} = \frac{1}{N}\sum\left(\frac{v_i - M_v}{\sigma_v}\right)^3 = \frac{1}{\sigma_v^3}\sum\frac{\bar{v}^3}{N} = \frac{\mu_{3:v}}{\sigma_v^3}$$

We see, therefore, that although the values of $\mu_{1:t}$ and $\mu_{2:t}$ are always 0 and 1 respectively, the value of $\mu_{3:t}$ will possess an abstract value depending, nevertheless, upon the variates themselves. The expression, $\mu_{3:t}$, is known as the coefficient of *skewness* and is denoted

by the symbol $\alpha_{3:v}$, i. e.

$$\alpha_{3:v} = \frac{\mu_{3:v}}{\sigma_v^3} = \frac{\mu_{3:v}}{\mu_{2:v}\,\sigma_v} \tag{12}$$

Summary of Section I. From the viewpoint of *Elementary Mathematical Statistics,* we characterize a series of variates by its

(a) number, N,
(b) mean, M_v ,
(c) standard deviation, σ_v , and
(d) skewness, $\alpha_{3:v}$

The *moments about the mean*, μ_{nv} , are introduced solely to facilitate the determination of σ_v and $\alpha_{3:v}$. Other moments, $\mu'_{n:v}$, are used to simplify the numerical calculation of the moments about the mean, $\mu_{n:v}$.

Verbally, we may state that the mean serves as a convenient average, and the standard deviation measures the concentration of the variates about their mean.

A thorough discussion of the significance of the coefficient of skewness must be slightly deferred. We may say at this time merely that the value of $\alpha_{3:v}$ depends obviously upon the value of $\mu_{3:v}$ and that a glance at the last column of Table I will lend weight to the statement that a positive or negative skewness indicates a weighted preponderance of those variates which are considerably greater than, or less than the mean, respectively.

Finally, the operations of mathematical statistics, and even certain comparisons in descriptive statistics, require that we introduce the notion of a standard variate, defined as follows:

$$t_i = \frac{v_i - M_v}{\sigma_v}$$

SECTION II.

INDIRECT METHOD OF OBTAINING ELEMENTARY FUNCTIONS

10 One of the fundamental theorems of moments states that if a constant be added to, or subtracted from each variate of a series, the moments computed about the mean for the revised series will be

identical with the corresponding moments of the original series. By way of a simple example:

The mean of the following five variates is 138, consequently the values of $\bar{v}$ are as given below:

i	v_i	$\bar{v}_i$
1	133	−5
2	142	4
3	138	0
4	141	3
5	136	−2
Total	690	0

If we subtract, say, 130 from each of the variates, then for the revised series x_1, x_2, x_3, x_4 and x_5,

i	$M_o = 130$ x_i	$\bar{x}_i = x_i - M_x$
1	3	− 5
2	12	4
3	8	0
4	11	3
5	6	− 2
Total	40	0

$$M_x = \frac{40}{5} = 8, \quad M_v = 130 + 8 = 138$$

The value subtracted, 130, is termed the ***provisional mean***, and in general is designated by the symbol, M_o. It follows, therefore, that

$$x_i = v_i - M_o \tag{13}$$

$$M_v = M_o + M_x \tag{14}$$

$$\mu'_{n:x} = \frac{\sum x^n}{N} \tag{15}$$

(16) $$\mu_{n \cdot v} = \mu_{n \cdot x}$$

It is understood that the functions of x are defined in precisely the same manner as corresponding functions of v, that is

$$M_x = \frac{\sum x}{N}$$

$$\bar{x}_i = x_i - M_x$$

$$\mu'_{n \cdot x} = \frac{\sum x^n}{N}$$

$$\mu_{n \cdot \bar{x}} = \frac{\sum \bar{x}^n}{N}$$

etc.

11. Formula (13) follows from definition, although (14)—seemingly self-evident—needs proof. Thus by (13)

$$\begin{aligned} v_1 &= M_o + x_1 \\ v_2 &= M_o + x_2 \\ v_3 &= M_o + x_3 \\ &\cdot \quad\ \cdot \quad\ \cdot \\ v_N &= M_o + x_N \\ \hline \sum v &= NM_o + \sum x \end{aligned}$$

Dividing both sides through by N yields, by definition,

$$M_v = M_o + M_x \qquad Q.\ E.\ D.$$

Formula (15) is proved by means of (13) and (14) as follows:

$$\begin{aligned} \bar{x}_i &= x_i - M_x && \text{(Definition)} \\ &= (v_i - M_o) - (M_v - M_o) && \text{(Formulae 13 and 14)} \\ &= v_i - M_v \\ &= \bar{v}_i && Q.\ E.\ D. \end{aligned}$$

Since

$$\mu_{n \cdot v} = \frac{\sum \bar{v}^n}{N} \quad \text{and} \quad \mu_{n \cdot x} = \frac{\sum \bar{x}^n}{N}$$

and we have just shown that always for corresponding values

$$\bar{V}_i = \bar{x}_i$$

the truth of (16) is apparent.

12. A comparison of tables III and I will reveal an advantage of the indirect over the direct method of calculation.

TABLE III

i	V_i	$M_0=50$ x_i	x_i^2	x_i^3
1	51	1	1	1
2	49	- 1	1	- 1
3	60	10	100	1000
4	53	3	9	27
5	48	- 2	4	- 8
6	51	1	1	1
7	42	- 8	64	-512
8	50	0	0	0
9	51	1	1	1
10	52	2	4	8
11	54	4	16	64
12	53	3	9	27
13	52	2	4	8
Total		16	214	616

$$M_x = \frac{16}{13}$$

$$\mu'_{2:x} = \frac{214}{13} \qquad \mu_{2:x} = \mu'_{2:x} - M_x^2 = \frac{2526}{13^2}$$

$$\mu'_{3:x} = \frac{616}{13} \qquad \mu_{3:x} = \mu'_{3:x} - 3\mu_{2:x}M_x + 2M_x^3 = -\frac{21240}{13^3}$$

$$\sigma_x = \sqrt{\frac{2526}{13^2}} = 3.78918$$

$$\alpha_{3:x} = \frac{\mu_{3:x}}{\sigma_x\,\mu_{2:x}} = -\frac{21240}{2526\sqrt{2526}} = -.167303$$

$$M_v = 50 + \frac{16}{13} = 51\ 3/13$$

$$\sigma_v = \sigma_x = 3.78918$$

$$\alpha_{3:v} = \alpha_{3:x} = -.167303$$

It will be observed that the values

$$\mu_{2:x} = \frac{2526}{13^2} \quad \text{and} \quad \mu_{3:x} = \frac{-21240}{13^3}$$

agree exactly with those of Table I, namely

$$\mu_{2:v} = \frac{2526}{169} \quad \text{and} \quad \mu_{3:v} = \frac{-21240}{2197}$$

The following will illustrate an important advantage of the indirect method of determining the moments, $\mu_{n:v}$. Let us suppose that after computing the values of M_v, σ_v and $\alpha_{3:v}$ for the 13 variates of Table I we desire to delete the 13th variate, $v_{13} = 52$, and compute the values of M, σ and α_3 for the remaining twelve variates.

By the direct method of Table I, the revision would be quite laborious, but by the indirect method of Table III, revisions are made easily, as follows:

$$N = 13 - 1 = 12, \ \sum x = 16 - 2 = 14, \ \sum x^2 = 214 - 4 = 210,$$
$$\sum x^3 = 616 - 8 = 608$$

Consequently

$$M_x = \frac{14}{12}$$

$$\mu'_{2:x} = \frac{210}{12} \quad \mu_{2:x} = \mu'_{2:x} - M_x^2 = \frac{581}{6^2}$$

$$\mu'_{3:x} = \frac{608}{12} \quad \mu_{3:x} = \mu'_{3:x} - 3\mu'_{2:x}M_x + 2M_x^3 = -\frac{1600}{6^3}$$

$$\sigma_x = \frac{1}{6}\sqrt{581} = 4.01732$$

$$\alpha_{3:x} = \frac{-1600}{581\sqrt{581}} = -.114250$$

$$M_v = 50 + \frac{14}{12} = 51\ 1/6$$

$$\sigma_v = \sigma_x = 4.01732$$

$$\alpha_{3:v} = \alpha_{3:x} = -.114250$$

13. In a word, revisions of series arising from

(a) increasing or decreasing the number of variates,
(b) combining two or more series, or
(c) correcting the original variates

together with the resulting smaller numbers that result by employing the indirect method, lead us ordinarily to avoid using the direct method of section I in computing the fundamental functions, ***mean, standard deviation*** and ***skewness.***

In practice, one continually faces the problem of revision. Thus, in business statistics, publications serving as sources of data frequently are obliged to present revisions for estimates made in previous issues. Moreover, monthly and annual endeavors to bring statistics up to date require the addition of variates to series. In problems arising in the field of psychology and education, it may develop after preliminary calculations have been made that one or more observations of the original series must be deleted due to the presence of factors such as unusual physical or mental impairment at the time of examination, cheating, etc. Again, we may desire to combine the statistics for several distinct intervals, for several classes, or for various schools of a city or state, etc.

In the numerical examples above, calculations were made in terms of fractions, rather than decimals, in order to emphasize the fact that the direct and indirect methods will yield identical results. Ordinarily, decimals are employed, and the results will consequently differ slightly.

Section III

Frequency Distributions

14. In dealing with *large* groups of quantitative data, the computation of the elementary statistical functions and an appreciation of the variation in the magnitudes of the series of measurements is greatly facilitated by systematically presenting the data in the form of a *frequency distribution.* Such a distribution may present in tabular form

(a) each *different* variate observed, and
(b) the number of times that each different variate was observed in the investigation.

It is evident at the very outset, therefore, that if a frequency distribution merely reproduces precisely the same data that might otherwise have been listed serially, the values of M, σ and α_3 computed from such a frequency distribution must correspond exactly with the values of M, σ and α_3 that would have been obtained by the serial method. This *serial method* has been considered in the two preceding sections.

15. As an illustration, suppose that we consider the complete table from which the 13 variates, used in earlier computations, were taken. Since, according to the regulations, 17,000 numbers were withdrawn, we shall have 170 groups of one hundred numbers each, consequently 170 variates. These are listed below.

We shall see that one can compute the fundamental functions from the frequency distribution more readily than from Table IV. Again, certain phenomena are apparent at a glance at Table V, though by no means evident from a short inspection of Table IV. Thus the *range* of the variates is immediately observed in Table V, and the degree of symmetry in the distribution can be guessed rather accurately by one accustomed to computing the coefficient of skewness from distributions.

TABLE IV

Number of even numbers in 170 samples of 100 numbers each.

U. S. Order of Call, 1918

51	42	49	53	49	46	47	51	57	48
49	51	55	50	46	53	46	47	46	54
60	59	42	42	58	43	53	49	54	53
53	46	47	50	55	50	48	47	44	51
48	57	49	52	57	56	45	64	37	58
51	53	51	49	39	54	51	56	44	41
42	46	50	56	42	54	50	45	47	58
50	52	53	55	52	48	50	53	45	48
51	55	47	45	55	51	47	54	48	46
52	60	52	53	49	52	46	62	43	48
54	50	51	50	50	53	44	54	51	45
53	47	44	48	55	45	55	45	55	50
52	55	54	56	42	49	45	55	45	55
44	37	44	53	52	50	51	47	56	44
54	56	50	53	49	52	60	48	50	51
56	45	50	51	53	44	47	54	46	54
42	44	49	43	57	46	48	48	49	48

The frequency distribution for Table IV may be obtained readily by means of the "cross-five" method as follows:

TABLE V

Frequency Distribution for Data of Table IV

v	Tabulation	f
37	𝍷𝍷	2
38		0
39	𝍷	1
40		0
41	𝍷	1
42	𝍸 𝍷𝍷	7
43	𝍷𝍷𝍷	3
44	𝍸 𝍷𝍷𝍷𝍷	9
45	𝍸 𝍸	10
46	𝍸 𝍸	10
47	𝍸 𝍸	10
48	𝍸 𝍸 𝍷𝍷	12
49	𝍸 𝍸 𝍷	11
50	𝍸 𝍸 𝍸	15
51	𝍸 𝍸 𝍷𝍷𝍷𝍷	14
52	𝍸 𝍷𝍷𝍷𝍷	9
53	𝍸 𝍸 𝍷𝍷𝍷𝍷	14
54	𝍸 𝍸 𝍷	11
55	𝍸 𝍸 𝍷	11
56	𝍸 𝍷𝍷	7
57	𝍷𝍷𝍷𝍷	4
58	𝍷𝍷𝍷	3
59	𝍷	1
60	𝍷𝍷𝍷	3
61		0
62	𝍷	1
63		0
64	𝍷	1
Total		170

16. The above type of distribution should be differentiated from others in which it has been found advantageous to combine the variates

into *classes* and likewise to group together the corresponding frequencies. A distribution of grades will serve to illustrate this second type of distribution.

TABLE VI

Distribution of Examination Grades of 168 Students

Class	Frequency
0- 10	0
11- 20	2
21- 30	3
31- 40	5
41- 50	7
51- 60	16
61- 70	39
71- 80	45
81- 90	41
91-100	10
Total	168

Such a table does not represent *exactly* the original data in which the grades were recorded for each student as an integral number of per cents; nevertheless, it gives a very good idea of the general form of the distribution and enables us to compute the fundamental functions with a considerable degree of accuracy.

17. *Discrete Variates.* The distribution of Table V is obviously one in which the variates can, from their very nature, be expressed only as integers. A distribution of this type is termed one of *discrete variates*, or one of a *discrete variable*. Common illustrations of this type are to be found in distributions of the number of individuals in a family, the number of petals on a flower, the number of coins turning up heads, etc.

18. *Continuous Variates.* In the majority of distributions the variates by their nature may differ by infinitesimals, and the observed values, as recorded, are merely more or less accurate estimates of the *true values*, which never can be established with ***absolute*** accuracy by any method of measurement. Thus the variates in the case of heights may be correct to the nearest inch, one-hundredth of an inch, or even the one millionth part of an inch, etc., but theoretically it can be shown that the chances that any measurement of a continuous variable is exact is about one in infinity. A frequency table for the distribution of continuous variates must always, therefore, be one of ***grouped frequencies.***

19. The fundamental differences between distributions which may be classified as

(a) discrete
(b) grouped discrete, and
(c) continuous

are of vital importance whenever the accurate determination of the mean, standard deviation, or skewness, is concerned. We shall now illustrate in detail and by numerical examples the procedure which should be followed in each case.

20. *Frequency Distributions of Discrete Variates.*

If 180 dice were thrown, and a throw of a six spot counted a success, then the *expected* frequencies of successes that would be obtained in one thousand such trials are as follows:

TABLE VII

V	f	$M_o = 30$ x	x^2	x^3
15	1	−15	225	−3375
16	1	−14	196	−2744
17	2	−13	169	−2197
18	4	−12	144	−1728
19	6	−11	121	−1331
20	10	−10	100	−1000
21	16	− 9	81	− 729
22	23	− 8	64	− 512
23	31	− 7	49	− 343
24	41	− 6	36	− 216
25	51	− 5	25	− 125
26	61	− 4	16	− 64
27	69	− 3	9	− 27
28	75	− 2	4	− 8
29	79	− 1	1	− 1
30	80	0	0	0
31	77	1	1	1
32	72	2	4	8
33	64	3	9	27
34	56	4	16	64
35	46	5	25	125
36	37	6	36	216
37	29	7	49	343
38	22	8	64	512
39	16	9	81	729
40	11	10	100	1000
41	8	11	121	1331
42	5	12	144	1728
43	3	13	169	2197
44	2	14	196	2744
45	1	15	225	3375
46	1	16	256	4096

$\sum f = 1000$ $\qquad M_o = 30$

$\sum xf = -27$ $\qquad M_x = -.027$

$\sum x^2 f = 24687$ $\qquad \mu'_{2:x} = 24.687$

$\sum x^3 f = 11259$ $\qquad \mu'_{3:x} = 11.259$

$\mu_{2:x} = 24.6863$ $\qquad \sigma_x = 4.96853$

$\mu_{3:x} = 13.2586$ $\qquad \sigma_x \mu_{2:x} = 122.655$

$$\alpha_{3:x} = .108097$$

$M_v = 29.973, \qquad \sigma_v = 4.96853, \qquad \alpha_{3:v} = .108097$

Explanation. Since this distribution of discrete variates is an exact reproduction of the original data listed serially, we know that the moments obtained by the frequency distribution method must be identical with those which would have resulted had the serial method been employed. In fact

$$(17) \qquad \begin{cases} \sum f = N, \\ \sum xf = \sum x, \\ \sum x^2 f = \sum x^2; \text{ and} \\ \sum x^3 f = \sum x^3. \end{cases}$$

Numerically, $\sum x^n f$ is absolutely equivalent to $\sum x^n$. However, $\sum x^n f$ implies more; it indicates a brief and systematic method of attaining a total in which multiplication replaces repeated additions. Thus, in the serial method the value $x = 5$ would be added 46 times during the numerical determination of $\sum x$. In the frequency distribution method one multiplication, 5 x 46, represents likewise the contribution of this variate to the total $\sum xf = \sum x$.

If a computing machine be not available, the headings of Table VII should be

v	f	x	xf	x^2f	x^3f

and the totals $\sum x^n f$ obtained by a detailed process. With the aid of a computing machine the values of $\sum x^n f$ may be obtained readily by a continuous process, and it is necessary to record only the totals.

Since

$$(x+1)^3 = x^3 + 3x^2 + 3x + 1$$

it follows that

$$(18) \quad \sum (x+1)^3 f = \sum x^3 f + 3\sum x^2 f + 3\sum xf + \sum f$$

Formula (18) is known as *Charlier's check.* By associating with each value, f the value of x^3 appearing on the next lower line, the value of $\sum (x+1)^3 f$ may be obtained as readily as that of $\sum x^3 f$. Then if equation (18) be satisfied we may assume with a considerate

degree of confidence that all five summations have been accurately determined.

It follows that we may now write, employing (17),

$$(19) \qquad \begin{cases} \mu'_{2 \cdot x} = \dfrac{\Sigma x^2 f}{\Sigma f} \\[2ex] \mu'_{3 \cdot x} = \dfrac{\Sigma x^3 f}{\Sigma f} \end{cases}$$

and observe that here, as in the serial method,

$$\mu_{2 \cdot x} = \mu'_{2 \cdot x} - M_x^2$$

$$\mu_{3 \cdot x} = \mu'_{3 \cdot x} - 3M_x \mu'_{2 \cdot x} + 2M_x^3$$

$$M_v = M_o + M_x$$

$$\mu_{2 \cdot v} = \mu_{2 \cdot x}$$

etc.

21 *The Grouping of Discrete Variates.* Occasionally frequency distributions of discrete variates contain so many different variates that some sort of grouping must be employed. Thus, the distribution of Table VII and the numerical calculations may be abbreviated as in Table VIII.

Explanation. The *class mark* of a class is defined as the arithmetic mean of the greatest and least variates that can occur within that class. In Table VII, we might have used the class marks as values of v, but the use of a provisional mean, as has already been demonstrated, saves a large amount of labor.

TABLE VIII (Unadjusted)

Class	Class Mark	f	$M_o=30,\ \lambda=3$ x
14-16	15	2	- 5
17-19	18	12	- 4
20-22	21	49	- 3
23-25	24	123	- 2
26-28	27	205	- 1
29-31	30	236	0
32-34	33	192	1
35-37	36	112	2
38-40	39	49	3
41-43	42	16	4
44-46	45	4	5

$\sum f = 1000$ $\qquad M_o = 30,\ \lambda = 3$

$\sum xf = -9$ $\qquad M_x = -.009$

$\sum x^2 f = 2817$ $\qquad \mu'_{2:x} = 2.817$

$\sum x^3 f = 405$ $\qquad \mu'_{3:x} = .405$

$\mu_{2:x} = 2.81692$ $\qquad \sigma_x = 1.67837$

$\mu_{3:x} = .481058$ $\qquad \sigma_x \mu_{2:x} = 4.72783$

$\alpha_{3:x} = .101750$

$M_v = 29.973,\qquad \sigma_v = 5.03511,\qquad \alpha_{3:v} = .101750$

The *class interval* is defined as the common difference between two consecutive class marks. In the example of Table VIII, the class interval has been chosen as the *unit* of x, consequently M_x and σ_x are expressed in class units. If λ denotes the class interval for a distribution, then

(20) $\qquad M_v = M_o - \lambda M_x$, and

(21) $\qquad \sigma_v = \lambda \sigma_x$

Thus in Table VIII we had

$$M_v = 30 + 3(-.009) = 29.973$$
$$\sigma_v = 3(1.67837) = 5.03511$$

Since the skewness is an abstract number, completely independent of the unit employed

(22) $$\alpha_{3:v} = \alpha_{3:x}$$

22. Table IX shows in the second, third, and fourth columns the values of M_v, σ_v and $\alpha_{3:v}$ which are obtained by various groupings of the data of Table VII. The grouping employed in Table VIII is listed as $D(3:2)$ in Table IX, the 3 denoting the number of different variates in each group, and the 2 designating the position of the first observed variate (i. e. $v = 15$) in the first grouping. Thus the classes of the grouping symbolized by $D(6:4)$ would be

12-17
18-23
24-29
etc.

From Table IX it may be observed that, although all of the values of M_v agree to a rather remarkable extent, nevertheless the unadjusted values of σ_v reveal the fact that an increase in the class interval is as a rule accompanied by an increase in the associated standard deviation and a decrease in the corresponding skewness.

23. In computing the moments $\mu'_{1:x}$, $\mu'_{2:x}$, and $\mu'_{3:x}$ for distributions of grouped frequencies, the assumption is made that each variate in a class may be treated as being numerically equal to the class mark. A mathematical investigation that lies beyond the scope of an elementary course shows that in the computation of M_x and $\mu_{3:x}$ it is entirely legitimate to treat each variate after this manner, but the demonstration also reveals that grouping tends to introduce a systematic error into the value of $\mu_{2:x}$. To eliminate this systematic tendency we find that one should introduce a correction and write

(23) $$\mu_{2:x} = \mu'_{2:x} - M_x - \frac{1 - 1/k^2}{12}$$

where k denotes the *number of different variates* that are grouped together in each class. Thus, in Table VIII we should have introduced as a correction

$$\frac{3^2 - 1}{12 \cdot 3^2} = \frac{2}{27} = .074074$$

TABLE IX

Comparison of Adjusted and Unadjusted Values of σ_v and $\alpha_{3:v}$

(1) Grouping	(2) M_v	(3) Unadjusted σ_v	(4) Unadjusted $\alpha_{3:v}$	(5) Adjusted σ_v	(6) Adjusted $\alpha_{3:v}$
$D(1:1)$	29.973	4.969	.108	4.969	.108
$D(2:1)$	29.972	4.992	.106	4.967	.108
$D(2:2)$	29.974	4.995	.107	4.970	.108
Avg. $D(2)$	29.973	4.935	.106	4.968	.108
$D(3:1)$	29.974	5.030	.109	4.963	.113
$D(3:2)$	29.973	5.035	.102	4.968	.106
$D(3:3)$	29.972	5.041	.101	4.974	.105
Avg. $D(3)$	29.973	5.035	.104	4.968	.108
$D(4:1)$	29.968	5.089	.096	4.964	.103
$D(4:2)$	29.976	5.094	.104	4.970	.112
$D(4:3)$	29.976	5.078	.104	4.970	.112
$D(4:4)$	29.972	5.096	.098	4.970	.105
Avg. $D(4)$	29.973	5.089	.100	4.968	.108
$D(5:1)$	29.975	5.160	.105	4.962	.118
$D(5:2)$	29.975	5.170	.097	4.972	.109
$D(5:3)$	29.970	5.167	.094	4.970	.105
$D(5:4)$	29.970	5.163	.085	4.966	.096
$D(5:5)$	29.975	5.170	.100	4.972	.112
Avg. $D(5)$	29.973	5.166	.096	4.968	.108
$D(6:1)$	29.974	5.247	.107	4.961	.126
$D(6:2)$	29.976	5.256	.099	4.971	.117
$D(6:3)$	29.972	5.259	.087	4.974	.102
$D(6:4)$	29.974	5.250	.085	4.965	.100
$D(6:5)$	29.970	5.251	.080	4.966	.094
$D(6:6)$	29.972	5.259	.091	4.974	.108
Avg. $D(6)$	29.973	5.254	.092	4.968	.108
$D(7:1)$	29.977	5.347	.097	4.959	.121
$D(7:2)$	29.971	5.347	.093	4.958	.117
$D(7:3)$	29.972	5.358	.087	4.971	.109
$D(7:4)$	29.966	5.354	.070	4.966	.087
$D(7:5)$	29.974	5.361	.088	4.974	.110
$D(7:6)$	29.975	5.360	.086	4.973	.108
$D(7:7)$	29.976	5.365	.084	4.978	.105
Avg. $D(7)$	29.973	5.356	.086	4.968	.108

This would have resulted in the following revision:

$$\mu_{2:x} = 2{,}74285 \qquad \sigma_x = 1.65616$$
$$\mu_{3:x} = .481058 \qquad \sigma_x \mu_{2:x} = 4.54260$$
$$\alpha_{3:x} = .105899$$

$$M_v = 29.973, \qquad \sigma_v = 4.96848, \qquad \alpha_{3:v} = .105899$$

Again, for $k = 7$ we would use

$$\mu_{2:x} = \mu'_{2:x} - M_x^2 - \frac{7^2 - 1}{12 \cdot 7^2} = \mu'_{2:x} - M_x^2 - \frac{4}{49}$$

When the simple adjustment of formula (24) is made, Table IX shows that the *systematic* errors in the values of σ_v and $\alpha_{3:v}$, caused by grouping, are eliminated. Thus in columns 5 and 6 the averages for each group are constant, consequently the errors remaining are accidental variations, which, due to a complete lack of compensation, still remain, but such discrepancies are not serious.

It should be noted that for distributions of discrete variates in which no grouping occurs, as in Table VII, the correction vanishes, since for $k = 1$

$$\frac{1 - 1/k^2}{12} = 0 \qquad (24)$$

24. *Frequency Distributions of Continuous Variates.* The following will serve as an illustration of the method of obtaining the fundamental functions for a distribution of continuous variates.

TABLE X

Weights of 1000 Female Students

(Original Measurements Made to Nearest 1/10 lb.)

Class (Pounds)	Class Mark $\lambda = 10$	f	$M_o = 114.95$ x
70- 79.9	74.95	2	–4
80- 89.9	84.95	16	–3
90- 99.9	94.95	82	–2
100-109.9	104.95	231	–1
110-119.9	114.95	248	0
120-129.9	124.95	196	1
130-139.9	134.95	122	2
140-149.9	144.95	63	3
150-159.9	154.95	23	4
160-169.9	164.95	5	5
170-179.9	174.95	7	6
180-189.9	184.95	1	7
190-199.9	194.95	2	8
200-209.9	204.95	1	9
210-219.9	214.95	1	10
Total		1000	

$\sum f = 1000$	$M_o = 114.95$
$\sum xf = 379$	$M_x = .379$ class units
$\sum x^2 f = 3089$	$\mu'_{2:x} = 3.089$
$\sum x^3 f = 8131$	$\mu'_{3:x} = 8.131$
$\mu_{2:x} = 2.86203$	$\sigma_x = 1.69175$
$\mu_{3:x} = 4.72769$	$\sigma_x \mu_{2:x} = 4.84184$

$\alpha_{3:x} = .976424$

$M_v = 118.74$ lbs., $\sigma_v = 16.9175^+$ lbs., $\alpha_{3:v} = .976424$

Explanation: The class mark has previously been defined as the mean of the greatest and least variates that can be included in a class. Since the original measuremtns were made to the nearest tenth of a pound, the *true limits* of the 150-159.9 class are 149.95-159.95, and

their mean is 154.95, which accordingly is the class mark in this instance. If the original measurements had been made to the *nearest pound*, then the classes would be written

150.0-159.0
160.0-169.0

and the true limits of the 150.0-159.0 class would be 149.5 and 159.5 pounds respectively, and the corresponding class mark would be 144.5 lbs. It is apparent, therefore, that a table of continuous variates should specify clearly the accuracy with which the original measurements were made, for the values of the class marks and consequently that of the mean, hinges on this point.

It will be noticed that in this example the class interval has again been taken as the unit of x, and this fact must be taken into consideration in determining the value of M_v and σ_v.

Since the assumption is also made that the class mark may represent the magnitudes of all variates occurring in that class, the question of correcting the second moment, $\mu_{2:x}$ again arises. Since in each class of a distribution of continuous variates an infinite number of different variates may occur, the correction is in this case

$$\frac{1 - 1/k^2}{12} = \frac{1}{12} .$$

Therefore, corresponding to formula (24), we must write, in order properly to adjust the second moment of a distribution of continuous variates

$$\mu_{2:x} = \mu'_{2:x} - M_x^2 - \frac{1}{12} \tag{25}$$

As before, neither the values of M_x nor $\mu_{3:x}$ require adjustment.

Summary of Section III. The frequency distribution is a device for presenting an extensive series of variates in a systematic and compact form. Not only are the phenomena of aggregation more readily perceptible by this method of presenting the data, but the calculations of the fundamental functions are facilitated.

The formulae for obtaining the mean, standard deviation and skewness are, with the exception of a single adjustment that may

arise, identical with those employed in the serial method. One need only observe that

$$N = \sum f$$
$$\sum x = \sum xf$$
$$\sum x^2 = \sum x^2 f$$
$$\sum x^3 = \sum x^3 f$$

The adjustment referred to is that we should in general regard

$$\mu_{2:x} = \mu'_{2:x} - M_x^2 - \frac{1 - 1/k^2}{12}$$

For ungrouped distributions of discrete variates this correction vanishes, since in this instance $k = 1$. For distributions of continuous variates, since here k would equal infinity, the correction is numerically equal to $1/12$.

These corrections will remove systematic errors in the standard deviation and skewness that arise from the phenomenon of grouping complete frequency distributions.

Editor's Note: This abstract of Elementary Mathematical Statistics will be continued in the May issue of the ANNALS.

BAYES' THEOREM [1]

By

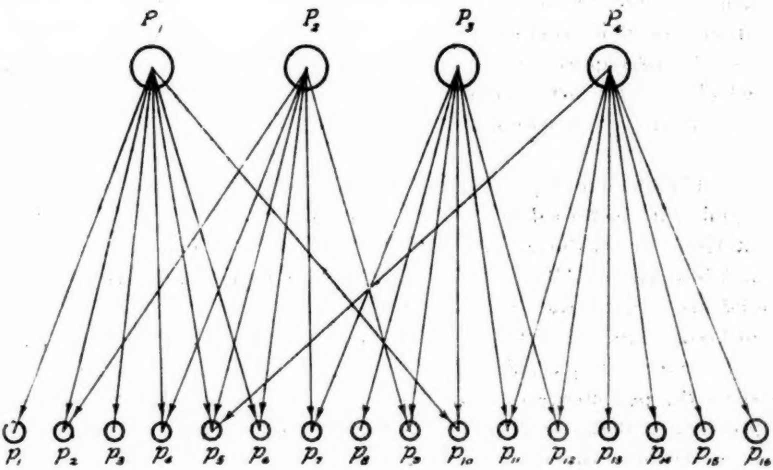
JOSEPH BERKSON

As for all established sciences, the typical problems of practical statistics have become inveterately attached to their several neat and convenient formulary solutions. To recall consideration of the basic reasoning underlying every-day statistical practice that applies to an elementary question may appear in the nature of an unnecessary disturbance of prevailing peace. If the experience of the writer is typical, however, vagueness or dubiousness of the premises inherent in a rule applied by rote will emerge to plague one in the conclusions, and a periodic return to fundamentals is as salutary for mental comfort as for the integrity of science itself. In what follows, an attempt will be made to go over the ground covered by Bayes' Theorem, and to point out its import for sound statistical reasoning. No claim is laid to mathematical originality at any specific points, but in the approach and synthesis will be found, we hope, a measure of instructive novelty.

A large class of statistical problems is typified in the following. A standard machine is known, from long experience, to produce a certain fraction P of imperfect products. What is the probability that in the next issue of n products, a fraction p will be imperfect?

We now present a related but not identical question. There is no available knowledge concerning the general practice of a machine; n products are examined and a fraction p found to be imperfect. What is the probability that the machine turns out generally a fraction P of imperfect products? The distinction between the two questions may be schematized as in Figure 1.

1. From the Department of Biometry and Vital Statistics of the School of Hygiene and Public Health (Paper No. 125); and the Institute for Biological Research of the Johns Hopkins University.



The values P_1, P_2; P_3; P_4 represent serially all the various fractions of imperfect products which might characterize particular machines, each one, let us say, determined by some definite combination of mechanical defects. Values p_1, p_2, p_3, etc., are the fluctuating fractions of imperfect products that might appear in the samples produced by these machines. Connected by arrows with P_1 are the randomly varying values of p that might result from P_1, with P_2 those that might result from P_2, etc., the weight of the arrows being proportional to the probability of the particular p concerned. It is to be noticed that each P may give rise to any of a number of p's and that some of the p's may result from any of a number of P's.

The first question in terms of the diagram is: "Given P_1, how probable is it that p_5 shall result?" The second is: "Given p_5, how probable is it that P_1 has been its source?" Answering the first, we calculate in the realm of the p's connected with P_1. In the second we calculate in the realm of the P's connected with p_5.

An answer to the first is given directly in terms of our every-day statistical reasoning. We say that the p's which result from P_1 can be adequately described as a normal distribution with $\sigma = \sqrt{\frac{P(1-P)}{n}}$, and from this the probability of any particular p calculated. The answer to the second is more difficult, and was given in general terms first by Bayes (1) in the theorem known by his name. Bayes' Theorem is not frequently used in applied statistics; yet the problems that

arise in practical situations would often seem to demand just such an answer as it provides. More often than not do we have a specific sample and inquire about the probable character of the universe from which it was drawn, in contra-distinction to the situation in which the universe is known, and the questions concern the possible samples.

The method of presenting the theorem here given will not follow rigidly any historical demonstration. Actually the calculation quantitatively of an "inverse probability" or the "probability of causes," was first given by Bayes. But he considered a purely geometric set-up and his solution was in terms of this conception. By implication he utilized a general principle first clearly stated later by Laplace, and furthermore, Laplace generalized the solution still more by arguing from the probability of a cause given by a particular sample, to the probability of the next sample. With this realized, then, that Bayes is to be credited with the original demonstration and Laplace for an important extension, we may proceed to a demonstration which is not exactly that of either.

I. *Problem.* We have an urn containing three balls. Each ball is colored black or white, and each color is equally likely. We draw one ball and it is black. What are the probable contents of the urn? We argue—the following are the possibilities:

I	II	III	IV
w w w	w w b	w b b	b b b

All of these possibilities, we say, are equally likely *a priori* and we have for the probabilities of the sample the following:

P_s I, the probability of a black sample from I = 0
P_s II, the probability of a black sample from II = 1/3
P_s III, the probability of a black sample from III = 2/3
P_s IV, the probability of a black sample from IV = 3/3

where P_s I is the probability of the sample s being drawn from urn I, P_s II from urn II, etc. We say now that the relative probabilities of the various urns are in proportion to the probabilities of the sample drawn, and we have

(a) $$P\,\text{I} : P\,\text{II} : P\,\text{III} : P\,\text{IV} = 0 : 1/3 : 2/3 : 3/3$$

where P I is the probability that, having drawn the ball, urn I was its source, P II that urn II was the source, etc.

Also, since the ball must have been drawn from some one of the urns, the total probability of one or another of the urns is unity and we have

(b) $$P\text{ I} + P\text{ II} + P\text{ III} + P\text{ IV} = 1$$

From (a) and (b) we have therefore

$$\begin{aligned} P\text{ I} &= 0 \\ P\text{ II} &= 1/6 \\ P\text{ III} &= 2/6 \\ P\text{ IV} &= 3/6 \end{aligned}$$

We now extend the problem to the case where the *a priori* probabilities of the various possible urns are not equal.

Suppose we say that there are many urns of the description I, II, III, IV in a large chamber, and that these are in proportion I : II : III : IV = 1 : 2 : 3 : 4. We now pick an urn at random and draw from it a ball, which turns out to be black. What is the probability that the urn is of some particular description? Proceeding as before, we have for the probabilities of the sample being drawn from the various urns the following:

$$\begin{aligned} p_s\text{ I} &= 1/10 \times 0 = 0 \quad \text{(Probability of urn} \times \text{probability of sample)} \\ p_s\text{ II} &= 2/10 \times 1/3 = 2/30 \\ p_s\text{ III} &= 3/10 \times 2/3 = 6/30 \\ p_s\text{ IV} &= 4/10 \times 3/3 = 12/30 \end{aligned}$$

where p_s I is the probability that such a sample s be drawn from urn I, etc.

And again on the principle that the probabilities of the urns are in proportion to the probabilities of the sample drawn, we have

$$P\text{ I} : P\text{ II} : P\text{ III} : P\text{ IV} = 0 : 2/30 : 6/30 : 12/30$$

and as preceding

$$P\,\mathrm{I} + P\,\mathrm{II} + P\,\mathrm{III} + P\,\mathrm{IV} = 1.$$

Therefore

$$\begin{aligned} P\,\mathrm{I} &= 0 \\ P\,\mathrm{II} &= 2/20 \\ P\,\mathrm{III} &= 6/20 \\ P\,\mathrm{IV} &= 12/20 \end{aligned}$$

We shall now generalize this solution.

Let π_1, π_2, π_3, etc. be the *a priori* probabilities of the various possible universes from which a sample is to be drawn. Let p_1, p_2, p_3, etc., be the probability of the sample being drawn from the respective universes. Then, a sample s having been drawn, the probability that its source is universe r is given by

$$P_r = \frac{\pi_r p_r}{\sum \pi p}$$

If all the universes are equally likely (our first case above), $\pi_1 = \pi_2 = \pi_3 = \pi_4$ and we have

(1) $$P_r = \frac{p_r}{\sum p}$$

If the equally probable universes are infinite in number, the P's varying by infinitesimal gradations from zero to unity, and p may assume any positive value less than 1, we may extend the last formula (1) by use of the calculus as follows:

Let x = any possible P between 0 and 1. From a universe x I draw a sample containing $r + s$ individuals, designated hereafter as a sample (r, s). The probability that it will contain r successes and s failures is given by

$$P_{(r,s)} = E_{r,s}\, x^r (1-x)^s$$

where $P_{(r,s)}$ is the probability that the sample (r, s) coefficient of the $(r+1)$th term in the Bernoulli expansion $= \frac{(r+s)!}{r!\,s!}$.

The probability of the sample of (r, s) coming from a universe

the P of which lies between x and $(x + dx)$ is therefore

$$_{x}^{x+dx}P_{(r,s)} = E_{r,s}\, x^{r}(1-x)^{s}dx \; .$$

where $_{x}^{x+dx}P_{(r,s)}$ is the probability that the sample (r, s) emanates from a universe whose P lies between x and $(x + dx)$. If the universe from which the sample is drawn may have a P anywhere between a and b, the probability of the sample (r, s) is

$$(2) \qquad _{a}^{b}P_{(r,s)} = E_{r,s} \int_{a}^{b} x^{r}(1-x)^{s}dx$$

and the probability that x is between a and b is therefore as in (1)

$$(3) \qquad _{a}^{b}P = \frac{\int_{a}^{b} x^{r}(1-x)^{s}dx}{\int_{0}^{1} x^{r}(1-x)^{s}dx}$$

where $_{a}^{b}P$ is the probability that the universe from which the sample (r, s) was drawn has a P between a and b. This is Bayes' Theorem in terms of the integral calculus.

Now, we ask the further question, what is the probability of a second sample containing m successes and n failures[1] being drawn?

If x be the p of the universe from which the sample (m, n) is drawn, and if P may vary from 0 to 1 we have analogously with (2)

$$(4) \qquad _{0}^{1}P_{(m,n)} = E_{m,n} \int_{0}^{1} x^{m}(1-x)^{n}dx$$

where $_{0}^{1}P_{(m,n)}$ is the probability that a sample (m, n) be drawn from universes whose P's vary between 0 and 1, and

$$E_{m,n} = \frac{(m+n)!}{m!\,n!}$$

1. Designated hereafter as the sample (m, n).

The probability of the event (m, n) occurring from any particular universe is given by the product of the probability of that universe and the probability of the event. The total probability of the event (m, n), i. e., the probability that the event (m, n) occurs at all from any universe, is, therefore, given by the product of form (3) with 0 and 1 substituted for a and b and (4), as follows:

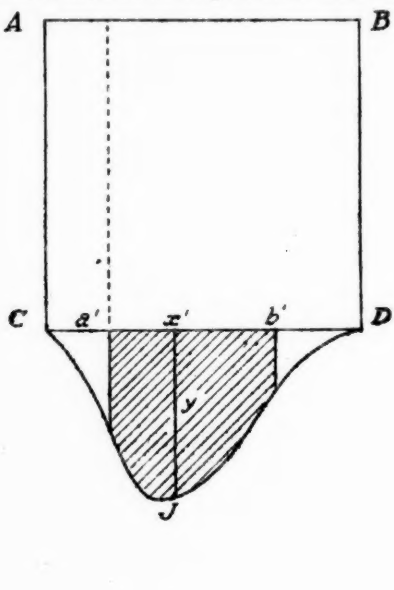
(5)
$$P_{(m,n):(r,s)} = \frac{(m+n)!}{n!\,m!} \frac{\int_0^1 x^{r+m}(1-x)^{s+n}\,dx}{\int_0^1 x^r(1-x)^s\,dx}$$

where $P_{(m,n):(r,s)}$ is the probability of a second sample (m, n) after a first sample (r, s) has been drawn.

This is Laplace's extension of Bayes' Theorem, somewhat modified.

Bayes' Solution.

It will be illuminating to derive this result by the method of Bayes. We shall follow his proof except to simplify his notation and to use the integral calculus where he used geometric demonstration.



ABCD is a square billiard table. A ball is thrown and comes to rest at a', through which a line is drawn parallel to AC. A second ball is thrown; if it stops to the left side of the line a', we designate a success, to the right, a failure. Before the first ball is thrown, what is the probability of the second ball succeeding r and failing s times in r plus s trials?

If the first ball comes to rest at x', the probability of a successful second throw is $\frac{Cx'}{CD} = p$, and of failure $\frac{Dx'}{CD} = q$. The probability of r successes and s failures with the first ball at x is then $\frac{(r+s)!}{r!s!} p^r q^s$.

Let us erect at each point x' along CD a distance y', so that

$$\frac{y'}{CD} = \frac{(r+s)!}{r!s!} p^r q^s \tag{6}$$

and connect the summits forming a figure as shown in Figure 2. At each point, of course, y' will be different because $p = \frac{Cx'}{CD}$, and $q = \frac{Dx'}{CD}$ will be different, but for any particular case, r and s remain constant.

The probability that the first ball shall fall between a and $(a+dx)$ is $\frac{dx'}{AD}$ and that the second ball shall therefore succeed r and fail s times is $\frac{y'}{CD}$. That both shall happen is therefore

$$\frac{y'}{CD} \times \frac{dx'}{CD}$$

and if x is to be between a' and b', the total probability is

$$\,_{a'}^{b'}P_{(r,s)} = \frac{1}{CD^2} \int_{a'}^{b'} y'dx'$$

where $\,_{a'}^{b'}P_{(r,s)}$ is the probability that the first ball fall between a' and b' and that a ball thrown subsequently $r+s$ times, succeed r and fail s times.

But $CD^2 =$ Area of AD and $\int_{a'}^{b'} y'dx' =$ Area of the shaded portion, $a'Jb'$. Therefore

(7) $$ {}_{a'}^{b'}P_{(r,s)} = \frac{Area\ a'Jb'}{Area\ AD} $$

The probability that the first ball fall between C and D and thereafter there occur r successes and s failures is similarly $\frac{Area\ CJD}{Area\ AD}$. But the first ball must fall somewhere between C and D; therefore the total probability of the second throws having r successes and s failures is given by

(8) $$ P_{(r,s)} = \frac{Area\ CJD}{Area\ AD} $$

With this established, the analysis proceeds.

Given the result of a series of throws to be r successes and s failures, what is the probability that the first ball has fallen between a' and b'? This we may obtain by the use of the solution already derived and the principle of compound probability[1].

Let x be the desired probability that the first ball fell between a' and b'. We have seen that the probability of r successes and s failures in the second series of throws is

$$ \frac{Area\ CJD}{Area\ AD} $$

from (8)

therefore the probability of the first falling between a' and b' and the experience (r, s) following is

$$ x \cdot \frac{Area\ CJD}{Area\ AD} $$

But we have shown that this combined probability is equal to

$$ \frac{Area\ a'Jb'}{Area\ AD} $$

from (7)

Therefore

(9) $$ x = \frac{Area\ a'Jb'}{Area\ CJD} $$

---

1. This step is very elaborately proved in Bayes' original paper by a circuitous demonstration.

This is Bayes' Theorem, as its author gives it. The additional part of his work is concerned with the quantitative estimate of the ratio.

We may now show that his solution is the same as that given in (3), as follows:

(10) $$y' = CD \times E_{r,s}\left(\frac{x'}{CD}\right)^r\left(1-\frac{x'}{CD}\right)^s$$ from (6)

where

$$x' = \text{distance from } C \text{ to } x'$$

$$E_{r,s} = \frac{(r+s)!}{r!s!}$$

Now

$$a' = a \times CD$$
$$b' = b \times CD$$

a and b having the meaning of equation (3). Assume the relationship

(11) $$x' = CD \times x$$

(12) $$dx' = CD \times dx$$

Then

$$\text{Area } aJb = \int_{x'=a'}^{x'=b'} y'dx'$$

$$= CD^2 \times E_{r,s} \int_{x=a}^{x=b} x^r(1-x)^s dx$$

(Substituting from (11) and (12)).

Similarly

$$\text{Area } CJD = CD^2 \times E_{r,s} \int_{x=0}^{x=1} x^r(1-x)^s dx$$

Therefore

$$Area\ \frac{aJb}{CJD} = \frac{\int_a^b x^r(1-x)^s\,dx}{\int_0^1 x^r(1-x)^s\,dx}$$

which is the same as formula (3) previously derived.

To be directly applicable to statistical problems formula (5) must be numerically evaluated. This is accomplished exactly for most practical instances only with a great amount of labor, and methods of approximation have been resorted to. For a few simple special cases the solution may be easily derived as follows:

An event has been tried N times with p successes and q failures. What is the probability that in the next single trial it will succeed?

Applying formula (5) to this instance, we have

$$r = p \qquad m = 1$$
$$s = q \qquad n = 0$$

and the desired probability is given by

$$P = \frac{\int_0^1 x^{p+1}(1-x)^q\,dx}{\int_0^1 x^p(1-x)^q\,dx}$$

Now

$$\int_0^1 x^a(1-x)^b\,dx = \frac{a!\,b!}{(a+b+1)!}$$

From which we have

$$P = \frac{m+1}{m+n+2} = \frac{m+1}{N+2}$$

So that if nothing is known concerning an event except that it has been tried three times and succeeded twice, the probability that it will

succeed in the next trial is 3/5, not 2/3 as the more usual procedure would indicate. Again, if an event has occurred a thousand times without a failure, and we know concerning it nothing except that fact, the probability that it will fail next instance is 1/1002. If an event has never been tried at all, the probability that it will succeed on the first trial is 1/2.

An event has been tried N times and succeeded each instance. What is the probability that in the next d trials it will again succeed each time? Here

$$r = N \qquad m = d$$

$$s = 0 \qquad n = 0$$

and the desired probability is given by

$$P = \frac{\int_0^1 x^{N+d}\,dx}{\int_0^1 x^{N}\,dx}$$

$$= \frac{N+1}{N+d+1}$$

From this we conclude that if an event has succeeded 25 times and never failed, the probability that in 25 further trials it will again not fail even once is 26/51, or in general if an event has never failed in N trials, the probability that N further trials will yield no failure is about 1/2.

Discussion.

To precisely what position in the methodology of applied statistics Bayes' Theorem will eventually become adjusted, it is impossible at this point in its development to say with certainty. The literature on the subject, as soon as it leaves the realm of purely hypothetical situations, is rife with disagreement, and clarification remains a contemporary problem. In this brief presentation, no attempt can be made to adequately summarize the various views concerning the questions at issue. We may, however, consider a few points that have disciplinary value for statistical thinking rather than any immediate practical utility.

It is basic to the aims of statistical calculations to estimate the

probability of given experiences from assumptions of pure random variation. A consideration of the logic involved in the development of Bayes' Theorem is useful in bringing out the inadequacy of the reasoning by which our most ordinary statistical procedures attempt to accomplish this. If, having observed a probability p, we estimate the standard deviation of succeeding samples of n by $\sqrt{\frac{pq}{n}}$, we imply tacitly that in the universe from which the sample was drawn, the chance of a success is the p of our observation. The reasoning leading to, and formula (3) itself, indicate how unwarranted this is. Our knowledge of the universe which generated the sample is never given with certainty by the sample. Indeed, formula (3) states a probability for any particular universe that may be assumed. With only a sample as the source of knowledge, and without Bayes' Theorem, we have no clue as to the nature of the generating universe. But, if we do not know the universe, how are we to calculate the character of its samples? One answer is to take refuge in formula (5), i. e. use Bayes' Theorem. As a practical solution of the difficulty this has two major objections: first, there are no existing tables for making the necessary calculations without prohibitive arithmetic labor; second, even if the evaluation could be effected there are reasons to doubt the validity of its application. For the formula in question rests on the assumption that all the probabilities from zero to unity which might characterize the universes from which we draw samples are *a priori* equally likely, the socalled assumption of the equal distribution of ignorance. Now this is an exceedingly questionable assumption, and it is partly on these grounds that Keynes rejects outright the possibility of applying probability to actual experience. It must be admitted, we think, that it is difficult to see what there is to justify the assumption that every sort of general universe from which arise the events of experience is equally likely. Would it not appear the more reasonable hypothesis that these universes are themselves "events," samples of some larger universe; and why should this be extremely different in the distribution of its probabilities from the universes that we ordinarily meet? There are writers, however, who, admitting that the assumption is to be questioned, believe it may be subjected to experimental test, and have essayed to actually sample at random the probabilities that characterize the universes of our experience. It would be impertinent to assert that an experimental investigation is bound to be futile, but the utility of this sort of procedure seems to us exceedingly dubious. We doubt indeed that any clear meaning can be assigned to the concept of "the universes of our experience," of which random samples are to be obtained. But granting the existence of such a

distribution of *a priori* probabilities we doubt the relevancy of its estimation to any practical problem. In any actual investigation, we deal with a definite slice of possible experience; an anthropologist is not concerned with the universes dealt with in the investigation of an economist or an epidemiologist. If *a priori* probabilities are of interest to him, they are those that obtain in his peculiar world of observation. It appears to us quite as wide of the mark aimed at, to call in a formula which obtains its *a priori* probability from experience in general, as to obtain it from the unique experience at hand, and indeed it may be argued that, as between the two, the latter is the more reasonable.

What then does all this come to? Does it mean that the entire structure of established statistical procedure rests on quicksand, to be toppled over by anyone armed with a reading of Bayes' Theorem? We are inclined to the belief held by Keynes that, so far as *logic* is concerned, this is substantially true. As regards this, however, it is at bottom in no worse plight than any current scientific procedure when its fundamental assumptions are hard pressed. But we do not rest the matter here. All this admits is that applied statistics, like all applied science, is not founded on unquestionable premises and invulnerable logic. It is perfectly consistent to add that *in general* its formulae are good *approximations*. How good? This is a question permitting no dogmatic comprehensive answer. Differently good for different situations. Some idea of the degree of approximation may be obtained for given assumed conditions by direct calculation. It may be shown, for instance, that under certain conditions results obtained by way of Bayes' Theorem or the more usual "normal" distribution render not very different results, and these conditions, indeed, approach the ones we most frequently encounter. But, in general, a more satisfactory answer is furnished in the pragmatic consideration that our formulae have in fact been widely used and experience has not violated their anticipations. This is the fact that we would stress, because it throws into relief the experimental as opposed to the mathematical foundation of statistics. Comforted on the one hand that experience in general supports our procedures, the considerations we have elicited in this discussion will emphasize equally their shifting approximation. The clear minded and careful worker will keep this constantly in mind and shun literal interpretation of conclusions drawn from formulae applied to extreme cases. No scientist worth his salt will permit himself the use of formulae the premises of which he has not examined. But the statistician, because of the great variability of

the data with which he is likely to deal, stands in special need of this precaution. Where statistics run counter to what appears to be the general experience, it is a wise rule to re-examine the statistics rather than to indict forthwith the dependability of the experience. Such an attitude would modify considerably much that is found in current statistical literature and it would modify it in the direction of greater soundness.

REFERENCES

1. Bayes, Thomas. Phil. Trans., 1763, LIII, 370; 1764, LIV, 296.
2. Coolidge, Julian L. Probability, Chapter VI.
3. DeMorgan, Augustus. An Essay on Probabilities, Chapter III.
4. Keynes, Maynard. Treatise on Probability, Chapter XXX.
5. Pearson, Egon. Biometrika, 1925, XVII, 388.
6. Pearson, Karl. Phil. Mag. 1927, 13, 365.
7. Todhunter, I. A History of the Mathematical Theory of Probability, Chapter XIV.
8. Wishart, John. Biometrika, 1927, XIX, 1.

A MATHEMATICAL THEORY OF SEASONALS

By

STATISTICAL DEPARTMENT, DETROIT EDISON CO.

The graph of any time series may be assumed to be a compound curve which is dependent upon the following factors:

Secular trend,	$f(x)$
Cycle,	$c(x)$
Seasonal	$s(x)$, and
Residual errors,	ϵ_x

If we designate the xth term of the observed time series by ${}_o y_x$, we have that

(1) $${}_o y_x = f(x) \cdot c(x) \cdot s(x) + \epsilon_x$$

It also follows that the standard error, based on our hypothesis, is

(2) $$\sigma_\epsilon = \sqrt{\frac{\sum \epsilon_x^2}{N}}$$

In making predictions, we desire that the standard error of estimate be a minimum, and this requires that $\sum \epsilon^2$ be also a minimum.

In dealing with data covering a period of years, i. e. $12\,n$ months, we observe that

$$\sum \epsilon^2 = [{}_0y_1 - f(1)\cdot c(1)\cdot s(1)]^2$$
$$+ [{}_0y_2 - f(2)\cdot c(2)\cdot s(2)]^2$$
$$\text{-----------------------}$$
$$+ [{}_0y_{12} - f(12)\cdot c(12)\cdot s(12)]^2$$
$$+ [{}_0y_{13} - f(13)\cdot c(13)\cdot s(1)]^2$$
$$+ [{}_0y_{14} - f(14)\cdot c(14)\cdot s(2)]^2$$
$$\text{-----------------------}$$
$$+ [{}_0y_{12n-11} - f(12n-11)\cdot c(12n-11)\cdot s(1)]$$
$$+ [{}_0y_{12n-10} - f(12n-10)\cdot c(12n-10)\cdot s(2)]$$
$$\text{-----------------------}$$
$$+ [{}_0y_{12n} - f(12n)\cdot c(12n)\cdot s(12)]$$

Let us now find the values of $s(1)$, $s(2)$, . . . $s(12)$ that will minimize the standard error of estimate. Placing the partial derivative of $\sum \epsilon^2$ with respect to $s(1)$ equal to zero, yields

$$\frac{\partial}{\partial s1} = 2[{}_0y_1 - f(1)\cdot c(1)\cdot s(1)][-f(1)c(1)]$$
$$+ 2[{}_0y_{13} - f(13)\cdot c(13)\cdot s(1)][-f(13)\cdot c(13)]$$
$$\text{-----------------------}$$
$$+2[{}_0y_{12n-11} - f(12n-11)\cdot c(12n-11)\cdot s(1)][-f(12n-11)\cdot c(12n-11)] = 0$$

Solving

$$s(1) = \frac{\sum^{(1)} {}_0y_x\, f(x)\cdot c(x)}{\sum^{(1)} f^2(x)\cdot c^2(x)}$$

where we understand that $\sum^{(1)} {}_o y_x \cdot f(x) \cdot c(x)$ means the sum of the products of ${}_o y_x$, $f(x)$ and $c(x)$ taken from the *first* month of each year, and similarly for $\sum^{(1)} f^2(x) \cdot c^2(x)$

The partial derivative with respect to $s(2)$ yiel

$$s(2) = \frac{\sum^{(2)} {}_o y_x \cdot f(x) \cdot c(x)}{\sum^{(2)} f^2(x) \cdot c^2(x)}$$

and in fact

$$s(i) = \frac{\sum^{(i)} {}_o y_x \cdot f(x) \cdot c(x)}{\sum^{(i)} f^2(x) \cdot c^2(x)} \tag{3}$$

Thus the seasonal for July is a function only of the various July values of the observed series, the secular trend and the cycle factors.

Since both $f(x)$ and $c(x)$ are smooth functions, it follows that their product, which we shall designate by $\psi(x)$, represents a smooth function which is merely that part of the time series which would remain if the accidental and seasonal fluctuations were eliminated. The formula for the seasonal index for the i th month may therefore be written

$$s(i) = \frac{\sum^{(i)} {}_o y_x \cdot \psi(x)}{\sum^{(i)} \psi^2(x)} \tag{4}$$

At this point we may recall the fact that in fitting a curve of the type $y = k\psi(x)$ to observed data by the *Method of Least Squares,*

$$k = \frac{\sum {}_o y_x \cdot \psi(x)}{\sum \psi^2(x)}$$

whereas if the *Method of Moments* be employed

$$k = \frac{\sum {}_o y_x}{\sum \psi(x)}$$

Experience in various statistical applications demonstrates that the two methods yield approximately the same results. Borrowing from this experience, we shall choose the simpler form and write in-

stead of formula (4)

$$s(i) = \frac{\sum_0^{(i)} y_x}{\sum^{(i)} \psi(x)} \tag{5}$$

So far as theoretical considerations are concerned (4) may be superior to (5), but the fact that the latter formula enables us to obtain seasonals by a method far simpler than would result by using formula (4), requires that we choose (5) in preference to (4). Ordinarily the difference in results obtained by using both formulae is less than one-half of one per cent.

Verbally, formula (5) states merely that *the seasonal index for any month is the ratio of the total of the variates for the month in question to the total that would have been experienced if neither accidental nor seasonal influences were present.*

We now are forced to find a simple method of obtaining values of $\sum^{(i)} \psi(x)$.

Let T_{i-3}, T_{i-2}, T_{i-1}, T_i, T_{i+1}, T_{i+2}, and T_{i+3} denote the total production for seven consecutive years. If we assume that the effect of both seasonal influences and accidental or residual fluctuations is to shift the production from one month to another, but nevertheless *to leave the total production for each year practically unchanged,* then a smooth curve passing over the seven year period, and preserving the annual totals, may be assumed to afford a representation of $\psi(x)$. We, therefore, determine the equation of a parabola of the sixth degree in such a manner that the areas under this curve for seven equidistant unit intervals are equal respectively to $T_{i-3}, T_{i-2}, \cdots T_0, \cdots T_{i+2}, T_{i+3}$. Fitting six degree parabolae to successive seven year intervals it is possible to deal with a time series of any length.

By adding together the interpolated values for all the January values of $\psi(x)$, and similarly for the other months, we can show that

$$\sum^{(i)} \psi(x) = c_{1:i}\, T_1 + c_{2:i}\, T_2 + c_{3:i}\, T_3 + c_{4:i} \left[T_4 + T_5 + \cdots T_{n-3} \right] + c_{5:i}\, T_{n-2} + c_{6:i}\, T_{n-1} + c_{7:i}\, T_n \tag{6}$$

where the values of the coefficients are as given in Table I.

In order to compare the efficiency of this method with another method of computing seasonals, it is necessary that each formula be tried out on some series for which the true values of the seasonal indices are known. We know in advance, of course, that there exist many satisfactory methods of obtaining seasonals, but we also desire to know something about the amount of time that each method requires as well as their relative accuracy.

The theoretical series, on which we shall try out two methods of computing seasonals, is built up from data taken from an article, "Statistical Analysis and Projection of Time Series," written and published by the statistical division of the American Telephone and Telegraph Company. After eliminating from the *Production of Pig Iron* series both trend and seasonal influences, the factors of Table II remained. We shall consider these, therefore, as the combination of "Cycle and residual" factors.

Although smoothing this data by a proper mathematical formula would eliminate the residual errors, nevertheless such procedure would introduce a bias in favor of the formula for computing seasonals that is proposed in this paper. The reason for this bias lies in the fact that most smoothing formulae are developed on the assumption that the smoothed ordinate lies on a parabola of a chosen degree, and since a similar assumption was made in our theory, it is evident that the proposed method will benefit most by employing a parabolic smoothing formula in obtaining the hypothetical cycle series.

For this reason the data of Table II, with additional data for one year on either side, was given to a draftsman with instructions to

(1) Plot the data of Table II

(2) draw free hand a *smooth* curve that to his mind best represented the general run of the data

(3) read off from his curve the approximate value of the smoothed statistics.

The data of Table III resulted.

In essential agreement with the American Telephone and Telegraph article, we shall assume a linear trend, the value for the first

month being 1511 and the monthly increment 8. The product of trend by cycle produces the theoretical values of $\psi(x)$ presented in Table IV.

TABLE I

Constants for computing seasonal indices

i	$c_{1:i}$	$c_{2:i}$	$c_{3:i}$	$c_{4:i}$	$c_{5:i}$	$c_{6:i}$	$c_{7:i}$
1	.12530	.07897	.08392	.083333	.08259	.08959	.03963
2	.11822	.07914	.08389	.083333	.08269	.08849	.04757
3	.11094	.07955	.08382	.083333	.08283	.08723	.05563
4	.10345	.08018	.08373	.083333	.08299	.08590	.06375
5	.09577	.08104	.08361	.083333	.08315	.08456	.07187
6	.08792	.08208	.08347	.083333	.08331	.08327	.07995
7	.07995	.08327	.08331	.083333	.08347	.08208	.08792
8	.07187	.08456	.08315	.083333	.08361	.08104	.09577
9	.06375	.08590	.08299	.083333	.08373	.08018	.10345
10	.05563	.08723	.08283	.083333	.08382	.07955	.11094
11	.04757	.08849	.08269	.083333	.08389	.07914	.11822
12	.03963	.08959	.08259	.083333	.08392	.07897	.12530

TABLE II

Cycle and Residual Series for Pig Iron Production

	1904	1905	1906	1907	1908	1909
January	–37.5	12.4	23.0	24.2	–43.3	– 7.9
February	–13.8	5.8	18.2	20.2	–36.4	– 7.8
March	–10.4	14.2	20.2	17.6	–40.8	–13.7
April	– .7	15.9	18.1	19.7	–42.0	–15.6
May	– 4.2	14.6	17.5	21.9	–43.0	–10.5
June	–15.2	10.4	15.0	23.1	–42.0	– 3.3
July	–27.4	7.0	16.8	23.9	–35.5	4.9
August	–25.3	11.1	9.8	21.6	–30.6	10.1
September	–12.5	15.4	12.7	19.0	–25.8	18.0
October	–12.1	18.6	20.2	21.4	–24.1	22.6
November	– 5.6	22.1	23.8	– 1.6	–19.0	24.2
December	1.1	20.7	24.9	–34.5	–12.1	27.0
	1910	**1911**	**1912**	**1913**	**1914**	**1915**
January	26.7	–17.7	– 8.0	19.5	–22.0	–35.9
February	21.5	–11.0	– 1.0	15.7	–16.7	–27.8
March	19.5	– 5.5	– .1	10.6	–10.5	–25.0
April	15.8	– 8.2	1.4	13.0	–10.8	–20.1
May	9.4	–17.9	5.4	13.9	–19.7	–16.3
June	8.2	–17.9	7.0	10.6	–21.9	– 6.9
July	2.5	–17.8	5.5	7.7	–20.3	.0
August	– 1.3	–13.6	8.1	5.1	–20.6	6.5
September	– 2.5	–10.1	7.1	4.7	–23.8	10.5
October	– 6.5	–10.3	11.0	.7	–33.6	15.1
November	–10.7	–10.5	12.8	– 7.8	–39.3	16.1
December	–18.1	– 9.8	17.8	–19.3	–40.6	21.0

TABLE III

Per Cent Cycle Series for Theoretical Distribution

	1904	1905	1906	1907	1908	1909
January	−38.2	6.6	15.5	16.3	−37.5	−12.2
February	−36.7	7.5	15.8	16.2	−38.1	− 7.1
March	−32.4	9.0	16.0	16.0	−39.1	− 2.4
April	−22.4	10.0	16.2	14.6	−39.4	2.1
May	−18.7	11.0	16.4	14.1	−39.4	5.0
June	−13.5	12.0	16.6	12.7	−38.5	10.0
July	− 9.8	13.1	16.6	11.6	−38.0	12.7
August	− 6.4	13.6	16.6	8.0	−36.7	15.6
September	− 3.7	14.1	16.7	5.0	−33.4	17.2
October	.0	14.5	16.6	.0	−29.1	18.4
November	2.3	14.8	16.5	−17.4	−23.8	19.6
December	4.4	15.0	16.5	−35.0	−17.0	20.0
	1910	**1911**	**1912**	**1913**	**1914**	**1915**
January	20.6	−12.3	− 7.2	10.8	−14.0	−32.3
February	20.6	−13.4	− 5.0	10.7	−17.0	−28.1
March	20.3	−13.6	− 3.0	10.6	−20.0	−23.6
April	19.5	−13.7	.0	10.0	−23.0	−18.1
May	17.9	−13.7	2.7	9.3	−24.7	−13.0
June	16.4	−13.6	4.8	8.5	−27.5	− 7.4
July	12.4	−13.4	6.9	7.1	−29.8	− 2.8
August	7.1	−12.8	7.9	6.0	−32.0	3.5
September	.0	−12.0	8.5	3.6	−33.5	9.0
October	− 6.6	−11.2	9.5	.0	−35.0	13.8
November	− 8.0	− 9.6	10.0	− 5.0	−36.0	15.7
December	−10.9	− 7.9	10.6	−10.0	−35.0	17.5

TABLE IV

Theoretical Trend and Cycle Series, $\psi(x)$

1904	1905	1906	1907	1908	1909
934	1713	1967	2092	1184	1748
962	1736	1981	2100	1178	1857
1032	1769	1994	2105	1164	1959
1191	1794	2007	2089	1163	2057
1254	1819	2020	2089	1168	2124
1342	1845	2032	2073	1190	2234
1406	1872	2042	2061	1205	2298
1467	1889	2051	2003	1235	2366
1517	1907	2062	1956	1305	2408
1583	1922	2070	1871	1395	2443
1628	1937	2077	1552	1505	2477
1669	1949	2087	1227	1646	2495
1910	**1911**	**1912**	**1913**	**1914**	**1915**
2517	1914	2115	2632	2125	1738
2527	1897	2173	2638	2058	1851
2530	1900	2226	2644	1990	1973
2523	1905	2303	2639	1921	2122
2498	1912	2373	2631	1885	2261
2476	1921	2430	2620	1820	2414
2400	1932	2488	2595	1768	2542
2295	1952	2519	2577	1718	2715
2151	1977	2542	2527	1686	2868
2017	2002	2574	2447	1653	3003
1994	2046	2595	2332	1633	3063
1938	2092	2618	2217	1663	3120

TABLE V

Theoretical Seasonal Factors

January	.99	May	1.04	September	.98
February	.93	June	.98	October	1.04
March	1.05	July	.98	November	.99
April	1.02	August	1.00	December	1.00

By multiplying the data of Table IV by the seasonals of Table V, a theoretical series would be obtained which would comprise the elements of trend, cycle and seasonal—lacking only chance or residual errors.

In order to obtain a series of chance factors that might serve as residual error factors, sixty cards were marked with integers totaling 1200. The cards were distributed, after shuffling, into twelve piles of five cards each, and the totals of each pile noted. These were taken as the residual factors for the first year, and the process was repeated for the following years. The chance factors of Table VI resulted.

Making allowance for residual errors as well as the seasonal factors, we obtain finally the theoretical series which we shall attempt to analyze, Table VII.

If the various methods of analyzing time series are sound, they should be able to break up this series into its elementary components—trend, cycle, seasonal and residual errors. A comparison of the results by different methods should indicate to some extent their respective merits. In attacking the ordinary observed time series by different methods and comparing results the difficulty is to tell, when all has been done, which of the methods is best. Unfortunately, if they disagree, we do not know which one is nearest the truth. Our theoretical series, however, enables us to compare results obtained by different methods, since we know the answers in advance, and also will serve students as a detailed example of time series synthesis.

TABLE VI

Residual Factors

	1904	1905	1906	1907	1908	1909
January	98	98	98	98	106	99
February	91	98	101	95	96	106
March	105	103	98	94	105	89
April	100	108	99	98	102	99
May	103	100	101	94	99	97
June	94	94	84	103	100	94
July	91	114	102	102	105	103
August	116	93	90	108	108	102
September	98	102	104	118	100	106
October	95	95	107	100	94	104
November	99	84	112	96	90	97
December	110	111	104	94	95	104
	1910	**1911**	**1912**	**1913**	**1914**	**1915**
January	96	102	98	98	97	98
February	107	99	102	95	93	89
March	91	95	97	92	105	92
April	110	96	106	104	104	83
May	104	109	98	104	110	108
June	102	92	108	105	97	113
July	94	92	109	91	91	103
August	98	103	98	98	96	98
September	100	102	104	96	100	105
October	105	103	94	112	102	107
November	95	107	93	108	98	102
December	98	100	93	97	107	102

TABLE VII

Theoretical Series

	1904	1905	1906	1907	1908	1909
January	906	1662	1908	2030	1242	1714
February	814	1582	1860	1855	1052	1831
March	1138	1913	2052	2077	1283	1831
April	1215	1976	2027	2088	1210	2077
May	1343	1892	2122	2043	1203	2143
June	1236	1700	1672	2093	1166	2058
July	1254	2092	2041	2060	1240	2320
August	1702	1757	1846	2163	1334	2413
September	1457	1906	2102	2262	1279	2502
October	1564	1899	2304	1946	1364	2643
November	1596	1611	2303	1475	1341	2378
December	1836	2163	2170	1153	1564	2595
Total	16061	22153	24407	23245	15278	26505

	1910	1911	1912	1913	1914	1915
January	2392	1933	2052	2554	2041	1687
February	2514	1746	2061	2330	1780	1532
March	2417	1895	2267	2554	2194	1906
April	2830	1865	2490	2800	2037	1796
May	2702	2167	2419	2845	2156	2539
June	2475	1732	2571	2696	1730	2674
July	2211	1742	2657	2314	1577	2566
August	2249	2011	2469	2525	1649	2661
September	2108	1976	2591	2377	1652	2952
October	2203	2144	2516	2850	1753	3342
November	1875	2168	2389	2494	1585	3093
December	1899	2092	2435	2150	1779	3182
Total	27875	23471	28917	30489	21933	29930

To obtain the values of the seasonal factors by means of formula (5) and Table I we need only observe that for the theoretical series

$$\begin{aligned} T_1 &= 16061 \\ T_2 &= 22153 \\ T_3 &= 24407 \\ T_4 + T_5 + \ldots T_9 &= 145291 \\ T_{10} &= 30489 \\ T_{11} &= 21933 \\ T_{12} &= 29930 \end{aligned}$$

Consequently we have

TABLE VIII

Seasonals by Interpolation Method

Month	$\Sigma_0 \, y_x$	$\Sigma \, \psi(x)$	s
January	22121	23587	.938
February	20957	23693	.885–
March	23527	23801	.988
April	24411	23911	1.021
May	25574	24023	1.065–
June	23803	24135	.986
July	24074	24246	.993
August	24779	24358	1.017
September	25164	24468	1.028
October	26528	24576	1.079
November	24308	24682	.985
December	25018	24785	1.009
Total	290264	290265	11.994

It is interesting to compare the seasonals of Table VIII with the corresponding set obtained by the method of "link relatives." The following table presents the series of link relatives for the theoretical series of Table VII.

TABLE IX

Link Relatives for the Series of Table VII

	1904	1905	1906	1907	1908	1909
January	.898	.952	.975	.914	.847	1.068
February	1.398	1.209	1.103	1.120	1.220	1.000
March	1.068	1.033	.988	1.005	.943	1.134
April	1.105	.957	1.047	.978	.994	1.032
May	.920	.899	.788	1.024	.969	.960
June	1.015	1.231	1.221	.984	1.063	1.127
July	1.357	.840	.904	1.050	1.076	1.040
August	.856	1.085	1.139	1.046	.959	1.037
September	1.073	.996	1.096	.860	1.066	1.056
October	1.020	.848	1.000	.758	.983	.900
November	1.150	1.343	.942	.782	1.166	1.091
December	.905	.882	.935	1.077	1.096	.922
	1910	**1911**	**1912**	**1913**	**1914**	**1915**
January	1.051	.903	1.004	.912	.872	.908
February	.961	1.085	1.100	1.096	1.233	1.244
March	1.171	.984	1.098	1.096	.928	.942
April	.955	1.162	.971	1.016	1.058	1.414
May	.916	.799	1.063	.948	.802	1.053
June	.893	1.006	1.033	.858	.912	.960
July	1.017	1.154	.929	1.091	1.046	1.037
August	.937	.983	1.049	.941	1.002	1.109
September	1.045	1.085	.971	1.199	1.061	1.132
October	.851	1.011	.950	.875	.904	.925
November	1.013	.965	1.019	.862	1.122	1.029
December	1.018	.981	1.049	.949	.948	

From the above we obtain the following:

TABLE X

Link Relative Seasonal Indices

Months	(1) Medians	(2) Chain Relatives	(3) (2) Adjusted	(4) Seasonal Indices
January	.913	100.0	100.0	97.5
February	1.112	91.3	91.3	89.0
March	1.019	101.5	101.4	98.8
April	1.024	103.5	103.3	100.7
May	.934	105.9	105.7	103.0
June	1.010	98.9	98.7	96.2
July	1.043	99.9	99.7	97.2
August	1.020	104.2	103.9	101.3
September	1.064	106.3	106.0	103.3
October	.914	113.1	112.7	109.9
November	1.024	103.4	103.0	100.4
December	.949	105.9	105.4	102.7
January		100.5	100.0	

The following exhibit of the results obtained by the two methods is interesting.

TABLE XI

Comparison of Interpolation and Link Relative Methods

Months	Actual Values	Interpolation Method		* Link Relative Method	
		Seasonal	Error	Seasonal	Error
January	.990	.938	-.052	.975	-.015
February	.930	.885	-.045	.890	-.040
March	1.050	.988	-.062	.988	-.062
April	1.020	1.021	.001	1.007	-.013
May	1.040	1.065	.025	1.030	-.010
June	.980	.986	.006	.962	-.018
July	.980	.993	.013	.972	-.008
August	1.000	1.017	.017	1.013	.013
September	.980	1.028	.048	1.033	.053
October	1.040	1.079	.039	1.099	.059
November	.990	.985	-.005	1.004	.014
December	1.000	1.009	.009	1.027	.027

The mean deviations and the standard deviations of the two methods show that both methods are about equally effective. This advantage of the interpolation method is scarcely worth mentioning. Nevertheless, the fact that the results are obtained with but a trivial amount of labor is important.

	Mean Deviation of Errors	Standard Deviation of Errors
Interpolation Method	.0269	.0337
Link Relative Method	.0277	.0338

STIELTJES INTEGRALS IN MATHEMATICAL STATISTICS

By

J. Shohat

(Jacques Chokhate)

Introduction. Stieltjes integrals, introduced into analysis in 1894-5[1], play an increasingly important role not only in pure mathematics, but also in theoretical physics and in the theory of probability. In mathematical statistics, however, their use, it seems, still remains very limited. And yet, one of the most remarkable features of Stieltjes integrals is that they represent, as the case may be, an integral proper or a sum of an finite or an infinite number of *discrete* aggregates. Thus *the statistician is enabled to treat in a single formula a continuous, as well as a discontinuous distribution.* This means far more than a mere simplification of writing. In fact, since Stieltjes integrals have many properties in common with Riemann and Lebesgue definite integrals, we can use all known resources of the theory of definite integrals (mean-value theorem, various inequalities), and therefore readily obtain general results which, otherwise, require special (often complicated) proofs. The advantage of such a treatment is particularly evident in the theory of interpolation, approximation, and mechanical quadratures.

Hence, the object of this paper is to present a general exposition of the properties and applications of Stieltjes integrals. Many of the results stated below are well known[2], and the proofs may be omitted. Some results are believed to be new (for example, extension of Tchebycheff and Hölder inequalities) and may prove useful in mathematical statistics. We close, as an illustration, with the theory of interpolation, for here, even in recently published books, the continuous and discontinuous cases are treated *separately* while the underlying ideas are *identical.*

---

1. Stieltjes: (a) Recherches sur les fractions continues, Oeuvres, v. II, p. 402-559; (b) Correspondence d'Hermite et de Stieltjes, v. II, p. 272, where these integrals are first mentioned in a letter (No. 351) to Hermite under date of October 25, 1892.

2. (a) Hobson, The Theory of Functions of a Real Variable, 2d. ed. (1921), v. I, p. 506-16, 605-09; (b) O. Perron, Die Lehre von den Kettenbrüchen (1913), p. 362-69.

I. *Definition and general properties.* Let $f(x)$ be continuous and $\psi(x)$ be bounded monotonic non-decreasing on the finite interval (a, b) $(a < b)$ Then, as is well known, the following limits exist:

$$\left.\begin{array}{l}\psi(x+0) = \lim\limits_{\epsilon\to 0} \left[\psi(x+\epsilon) - \psi(x)\right] \\ \psi(x-0) - \lim\limits_{\epsilon\to 0} \left[\psi(x-\epsilon) - \psi(x)\right]\end{array}\right. \quad (a \leq x \leq b)$$

If x is a point of discontinuity of $\psi(x)$, $\psi(x+0) - \psi(x-0)$ (>0) is called "saltus" of $\psi(x)$ at this point. The number of such points is at most denumerably infinite; the points of continuity of $\psi(x)$ are, therefore, everywhere dense in (a, b) $\psi(x)$ is R- integrable, and so is $\psi(x)x^k$ $(k = 0, 1 \cdots\cdots)$. The Riemann-Stieltjes integral (of $f(x)$ with respect to $\psi(x)$ $\int_a^b f(x)d\psi(x)$) is defined as follows:

(S)

$$\int_a^b f(x)d\,\psi(x) = \lim_{\substack{n\to\infty \\ max(x_{i+1}-x_i)\to 0}} \sum_{i=0}^{n-1} f(\xi_i)\left[\psi(x_{i+1}) - \psi(x_i)\right]$$

$$a = x_0 < x_1 < x_2 < \cdots\cdots < x_{n-1} < x_n = b$$

$$x_i \leq \xi_i \leq x_{i+1} \; (i = 0, 1, \cdots n-1)$$

The existence of the right-hand limit can be easily established. The continuity of $f(x)$ is here sufficient, but not necessary[1].

In many phases of mathematical statistics the case of a continuous $f(x)$ is evidently the more important, although many problems arising in the theory of probability require applications of the discontinuous case.

From the very definition (S) one may obtain many properties of Stieltjes integrals in common with the ordinary definite integrals. Thus:

(1) $\int_a^b d\,\psi(x) = \psi(b) - \psi(a)$

1. (a) Hobson, 1-c; (b) T. Hildebrandt, On Integrals Related to and Extension of the Lebesque Integrals, Bulletin of the American Mathematical Society (2), V. 24 (1918), p. 177-202; (c) Lebesque, Leçon sur l'intégration, 2d ed. (1928), p. 252-313.

(2) $\int_a^c f\,d\psi + \int_c^b f\,d\psi = \int_a^b f\,d\psi \qquad (a < c < b)$

(3) $\int_a^b (f_1 \pm f_2)\,d\psi = \int_a^b f_1\,d\psi \pm \int_a^b f_2\,d\psi$

(4) $\int_a^b Af\,d\psi = A\int_a^b f\,d\psi \qquad (A = \text{Const.})$

(5) $\left|\int_a^b f\,d\psi\right| \leq \int_a^b |f|\,d\psi$

(6) $\int_a^b f\,d\psi = f(\xi)\int_a^b d\psi \qquad (a \leq \xi \leq b$; mean-value theorem)

(7) $\int_a^b f_1\,d\psi \leq \int_a^b f_2\,d\psi$, if $f_1(x) \leq f_2(x)$ for $a \leq x \leq b$

(8) $\int_a^b \sum_{i=1}^{\infty} f_i\,d\psi = \sum_{i=1}^{\infty} \int_a^b f_i\,d\psi$

if $\sum_{i=1}^{\infty} f_i(x)$ converges uniformly in (a, b)

(9) $\int_a^b f\,d\psi = f\psi\Big|_a^b - \int_a^b \psi\,df \qquad$ (integration by parts)

(10) $\int_a^b f\,d\psi = \int_a^b f(x)p(x)\,dx$, if $\psi(x) = \int \phi(x)\,dx + c$

with $p(x) \geq 0$ in (a, b).

(10-bis) $\int_a^b f\,d\psi = \int_a^b f(x)\psi'(x)\,dx$, if $\psi'(x)$ exists

and is R- integrable in (a, b).

Let $\psi(x)$ have only a *finite* number of points of increase in (a, b).

$$(a = x_0 <)\, x_1 < x_2 < \cdots\cdots < x_n\, (< x_{n+1} = b)$$

with the saltus σ_i at $x = x_i$ $(i = 1, 2 \cdots\cdots n)$, so that $\psi(x)$ remains constant $= \sum_{s=1}^{i} \sigma_s$ for $x_i \leq x < x_{i+1}$ and $\psi(b) \sum_{i=1}^{n} \sigma_i$. Such functions, called *stepweise* functions ("fonction en escalier"), prove

very useful. Here

(11) $\int_a^b f\,d\psi = \sum_{i=1}^{n} \sigma_i f(x_i) \qquad \{\sigma_i = f(x_i+0) - f(x_i-0)\}$

If the number of points of increase is infinite

$$(a<)\, x_1 < x_2 < \cdots < x_n < \cdots, \quad \lim_{n\to\infty} x_n = b$$

(12) $\int_a^b f\,d\psi = \sum_{i=1}^{\infty} \sigma_i f(x_i)$.

Conversely, any sum $\sum_{i=1}^{n} u_i v_i$ can be represented as a Stieltjes integral in infinitely many ways. Let us introduce n positive numbers $\sigma_1, \sigma_2, \cdots \sigma_n$ a certain interval (a, b), n points $(a<)\, x_1 < \cdots < x_n (<b)$ (the choice of x_i σ_i depends upon the nature of the problem involved), and a stepwise function $\psi(x)$ having at $x = x_i$ a saltus σ_i ($i = 1, 2, 3 \cdots n$). Then, writing $u_i = \sigma_i w_i$, we may consider v_i, w_i as values taken respectively by some functions $f(x), \phi(x)$ at $x = x_i$ ($i = 1, 2, \cdots\cdots n$) . Hence,

(13) $\sum_{i=1}^{n} u_i v_i = \int_a^b f(x)\phi(x)\,d\psi(x)$

Formulae (11-13) show clearly the use of Stieltjes integrals for the representation of sums of discrete aggregates.

(14) $\int_a^b f\,d\psi \geq 0, \quad if\ f(x) \geq 0 \quad in\ (a, b)$

Here "=" takes place if and only if $\psi(x)$ has a finite or denumerably infinite number of points of increase in (a, b) (not everywhere dense) and $f(x)$ vanishes at all these points, for we exclude, of course, functions $f(x)$ which vanish at all points of continuity of $\psi(x)$ and therefore vanish identically in (a, b). *If $\psi(x)$ has infinitely many points of increase, while $f(x)$ vanishes in (a, b) only a finite number of times, without changing sign, then $\int_a^b f(x)\,d\psi(x) \neq 0$ and has the sign of $f(x)$.*

(15) $\int_a^b f(x)\, x^k\, d\psi(x) = 0 \quad (k = 0, 1, \cdots n-1)$ implies: $f(x)$ has at least n distinct roots inside (a, b) assuming that $\psi(x)$ has at least n points of increase[1].

---

1. This is a form of a theorem due to Perron (l-c. p. 368-69). If the number of such points is $m < n$. (15) shows only that $f(x)$ vanishes at all such points.

(16) $\int_a^b x^k d\psi(x) = 0 \quad (k = 0, 1, \cdots\cdots)$ implies:

$\psi(x)$ constant for $(a \leq x \leq b)$[1].

Since in the defiition (S) only the differences $\psi(x_{i+1}) - \psi(x_i)$ enter, it follows that a Stieltjes integral does not change its value if we replace $\psi(x)$ by $\psi(x) + c$. More precisely:

(17) $\int_a^b f d\psi_1 = \int_a^b f d\psi_2$

if the two monotonic non-decreasing functions $\psi_{1,2}(x)$ differ by an additive constant only at all points of continuity. Applying the mean-value theorem to $\int_x^{x+h} f(x) d\psi(x)$, we conclude:

(18) $F(x) = \int_a^x f(t) d\psi(t)$ is continuous at all points of continuity of $\psi(x)$ and therefore, almost everywhere in (a, b).

(19) $\lim_{h \to 0} \frac{F(x+h) - F(x)}{\psi(x+h) - \psi(x)} = f(x) \qquad (a \leq x \leq b)$

(20) $F'(x) = f(x)\,\psi'(x)$ at any point x, in (a, b), where $\psi'(x)$ exists.

One recognizes in (18-20) a generalization of the properties of the ordinary definite integral which is a special case, for $\psi(x) \equiv x$.

(21) $\phi(t) = \int_a^b f(x, t) d\psi(x)$ is continuous in $t\,(t_0 \leq t \leq t_1)$

if $f(x, t)$, continuous in x, is uniformly continuous with respect to $t\,(t_0 \leq t \leq t_1)$ for all values of x in (a, b). Moreover,

(22) $\frac{d\phi(t)}{dt} = \int_a^b \frac{\partial f(x, t)}{\partial t} d\psi(x)$

if $\frac{\partial f(x, t)}{\partial t}$ exists and is continuous in x and uniformly continuous in $t\,(a \leq x \leq b;\ t_0 \leq t \leq t_1)$.

1. If $\psi(x)$ has a finite number, n, of points of increase, then n such relations imply the same conclusion.

Notes. (i) The above results hold, with proper limitations and modifications, if $\psi(x)$ be of *bounded variation* in (a, b), for such a function can be represented as a difference of two monotonic non-decreasing functions $\psi_{1,2}(x)$ and we define in accordance with (S),

$$\int_a^b f\,d\psi = \int_a^b f\,d\psi_1 - \int_a^b f\,d\psi_2 \ .$$

(ii) In applications to probability and mathematical statistics $\psi(x)$ stands for the "cumulative law of distribution," so that

(23) $\psi(x)$ is monotonic non-decreasing from $\psi(a)=0$ to $\psi(b)=1$.

(24) For $(a \leq c < d \leq b)$ the integral $\int_c^d d\,\psi(x)$ = probability $P[c \leq x \leq d]$; $\int_a^b d\,\psi(x) = 1$.

(25) $\int_a^b f(x) d\,\psi(x) = E(f)$, i. e., the expected value or mathematical expectation of $f(x)$.

Let $w(x)$ $f(x)$ be continuous in (a, b), and $\alpha(x)$ be of bounded variation. Then,

(26) $\psi(x) = \int_a^x w(x) d\alpha(x)$ is of bounded variation,[1]

$$\int_a^b f(x) d\,\psi(x) = \int_a^b f(x) w(x) d\alpha(x)$$

Given an infinite sequence of functions $\psi_n(x)$ $(n = 1, 2, \cdots)$ of bounded variation in (a, b). If the total variation in (a, b) of all $\psi_n(x)$ does not exceed a fixed quantity M independent of n, and if, in addition, $\lim\limits_{n\to\infty} \psi_n(x) = \psi(x)$ exists for $a \leq x \leq b$, then[2]

(27) $\lim\limits_{n\to\infty} \int_a^b f(x) d\,\psi_n(x) = \int_a^b f(x) d\,\psi(x)$, for any continuous $f(x)$.

Notes. (i) (27) holds true if we know that $\lim\limits_{n\to\infty} \psi_n(x) = \psi(x)$ exists at all points of continuity of the sequence $\psi_n(x)$ and at $x = a, b$.

1. T. Carleman Leçons sur les équations intégrales singulières noyau réel et symétrique (Uppsala) (1923), p. 11-12.
2. Page 9 of preceding reference.

(ii) In applications to probability and statistics (27) is of great importance. In fact, consider $\psi_n(x)$ *as a sequence of variable laws of distribution approaching, as a limit, a certain fixed law of distribution* $\psi(x)$. Then(by (23), the total variation of any $\psi_n(x)$ in (a, b) is 1; (27) thus becomes applicable and shows that *under the said conditions the expected value of any continuous function in the variable law of distribution approached, as* $n \to \infty$ *its expected value in the limiting law of distribution.*

II. *Stieltjes Integrals Over an Infinite Interval.* We define

$$(28) \qquad \int_a^\infty f\,d\psi = \lim_{x\to\infty} \int_a^x f\,d\psi\ ; \qquad \int_{-\infty}^\infty f\,d\psi = \lim_{\substack{a\to-\infty \\ b\to\infty}} \int_a^b f\,d\psi$$

(similarly $\int_{-\infty}^b$), provided the right-hand limits exist as finite numbers. It is assumed that $\int_a^x f d\psi$, $\int_a^b f d\psi$ exist respectively for any finite $x > a$, and for any finite interval (a, b). For the existence of (28) it is necessary and sufficient that

$$(29) \qquad \left| \int_{x_1}^{x_2} f\,d\psi \right| < \epsilon \quad \text{for } x \geq \text{ a certain number } x(\epsilon),$$

$\epsilon > 0$ — arbitrarily small.

One sees readily that

(30) $\int_{-\infty}^{\infty} f\,d\psi$ exists, if $\int_{-\infty}^{\infty} d\psi$ does, and if $f(x)$ is bounded for all real values of x. The first of these conditions is satisfied if $\psi(x)$ is a law of distribution. We notice that any $\int_a^b f\,d\psi$ can be written as $\int_{-\infty}^{\infty} f\,d\psi$, if we agree to take $\psi(x) = \psi(a)$, $\psi(b)$ respectively for $x \leq a$, $\geq b$.

The formulae given above hold, in general, for infinite limits as well, *with the exception of those which require a double limiting process*, like 8, 21, 27, etc., where ordinarily additional precautions must be taken in the form of certain assumptions specifying the behaviour of $\psi(x)$ and of other functions involved at infinity. Thus, (8) is not valid in general for $(a, b) = (-\infty, \infty)$, and requires a more detailed discussion. Furmulae 21, 22 hold true if we assume, for example, the uniform boundedness and continuity with respect to t of the functions involved for all x in $(-\infty, \infty)$, and also the existence of $\int_{-\infty}^{\infty} d\psi(x)$, i. e. definite values for $\psi(\pm\infty)$.

Formula 17 deserves *special attention*: *in general, it is not true*

for an infinite interval, as was shown by Stieltjes[1].

III. *Approximate Evaluation of Stieltjes Integrals.* In practice, as in statistical computations, we evaluate $\int_a^b f\,d\psi$ approximately, replacing it by the right-hand member of (S), for a certain chosen n. The question arises regarding the error r_n of such an approximation. Let $\omega(a)$ represent the modulus of continuity of $f(x)$, i e.

$$(31)\qquad |f(x)-f(y)| \leq \omega(\delta)\ for\ |x-y| \leq \delta\ (a \leq x,\ y \leq b)$$

Then, if $x_{i+1} - x_i < h$ in (S) for $i = 0, 1, \cdots\cdots n-1$, we have

$$r_n = \int_a^b f\,d\psi - \sum_{i=0}^{n-1} f(\xi_i)\left[\psi(x_{i+1}) - \psi(x_i)\right]$$

$$= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} \left[f(x) - f(\xi_i)\right] d\psi(x)$$

$$(32)\qquad |r_n| \leq \omega(h) \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} d\psi = \omega(h)\left[\psi(b) - \psi(a)\right]$$

$$\left\{h = max.\ (x_{i+1} - x_i)\ ;\ i = 0, 1, \cdots\cdots\cdots n-1\right\}$$

(32) *answers the above question for any continuous* $f(x)$.

Special Case: Lipschitz condition[2]:

$$(33)\qquad |f(x)-f(y)| \leq \lambda\,|x-y| \qquad (a \leq x,\ y \leq b\ ;$$

$$|r_n| \leq \lambda h \left[\psi(b) - \psi(a)\right] \qquad \lambda = const.)$$

In (32, 33) *we replace* h *by* $h/2$, *if* (ξ_i) *in* (S) *is, as usual, the mid-point of the interval* (x_i, x_{i+1}) $(i = 0, 1, \cdots\cdots n-1)$.

It must be noticed, however, that the above considerations are

---

1. (l. c. p. 73, p. 505-06. (17) is closely related to the so-called "Moments-problem": find a monotonic non-decreasing function $\psi(x)$ in (a, b) with infinitely many points of increase, if all its moments $\gamma_k = \int_a^b x^k d\psi(x)$ $[k = 0, 1 \cdots]$ are given. *This problem, for* (a, b) *infinite, may be "indetermined," i. e. it may admit infinitely many solutions, while it is always "determined," for finite* (a, b). Stieltjes gives the following example:
$$\int_0^\infty x^k [1 + \lambda \sin (x^{1/4})] e^{-x^{1/4}} dx = \int_0^\infty x^k e^{-x^{1/4}} dx$$
$[\lambda$ ∴ constant, $k = 0, 1, \cdots\cdots]$, and $\psi(x) = 0 = \int_0^x [1 + \lambda \sin x^{1/4}] e^{x^{1/4}}$ is monotonic non-decreasing in $0, \infty$, if $|\lambda| \leq 1$.

2. If $f'(x)$ exists for $a \leq x \leq b$, then $\lambda^{-\infty}$ can be taken equal to max. $|f'(x)|$ *in* (a, b). If $f(x)$ is given graphically, λ can be found roughly as the maximum of the absolute va[illegible] .ne slope in (a, b).

not workable in general on an infinite interval, for here, in place of (31), we ordinarily have the more complicated relation

$$|f(x)-f(y)| \le \omega(x,y,\delta) \qquad (|x-y| \le \delta)$$

where $\omega(x,y,\delta) \to \infty$ with x, y (ex.: $f(x)=x^2$). Thus here, in order to obtain an inequality for the error, we must add to the right member of (32), where a, b are finite numbers properly chosen, two more terms—the upper limits of $\left|\int_{-\infty}^{a} f\, d\psi\right|$ and $\left|\int_{b}^{\infty} f\, d\psi\right|$, which we obtain by means of a suitable hypothesis concerning the behavior of $f(x)$, $\psi(x)$ at infinity.

IV. *Tchebycheff and Hölder Inequalities for Stieltjes Integrals*[1]. Hereafter $\psi(x)$ stands for a monotonic non-decreasing function defined on a certain interval (a, b), finite or infinite. Let $f_i(x)$, $\phi_i(x)\,[i=1,2,\cdots n]$ be continuous on (a, b)[2]. Then we have the following *fundamental transformation*:

$$(34)\qquad \begin{vmatrix} \int_a^b f_1\phi_1\,d\psi & \int_a^b f_1\phi_2\,d\psi & \cdots & \int_a^b f_1\phi_n\,d\psi \\ \int_a^b f_2\phi_1\,d\psi & \cdots & \cdots & \int_a^b f_2\phi_n\,d\psi \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \int_a^b f_n\phi_1\,d\psi & \cdots & \cdots & \int_a^b f_n\phi_n\,d\psi \end{vmatrix}$$

$$= \frac{1}{n!}\underbrace{\int_a^b\cdots\int_a^b}_{(n\ \text{times})} \begin{vmatrix} f_1(x_1) & \cdots & f_1(x_n) \\ \vdots & & \vdots \\ f_n(x_1) & \cdots & f_n(x_n)\end{vmatrix} \cdot \begin{vmatrix} \phi_1(x_1) & \cdots & \phi_1(x_n) \\ \vdots & & \vdots \\ \phi_n(x_1) & \cdots & \phi_n(x_n)\end{vmatrix} \prod_{i=1}^{n} d\psi(x_i).$$

The proof is very simple for $n=2$, for we can write

$$\int_a^b u(x)\,d\psi(x)\cdot\int_a^b v(x)\,d\psi(x)$$

$$\text{as} \int_a^b\int_a^b u(x_1)\,v(x_2)\,d\psi(x_1)\,d\psi(x_2)$$

and it may readily be extended to any n. Formula (34) yields many

1. Cf. my Note: Jacques Chokhate. Sur les intégrales de Stieltjes. Comptes Rendus, v. 189 (1929), p. 618-20.

2. In case $\psi(x)$ has a finite number of points of increase in (a, b), we require only definite values of all $f_i(x)$, $\phi_i(x)$ at these points.

interesting results by a proper choice of n, f_i, ϕ_i.

Examples: (i) $n=2$; $f_1\equiv\phi_1$, $f_2\equiv\phi_2$

(35) $$\int_a^b f_1^2 d\psi \int_a^b f_2^2 d\psi - \left(\int_a^b f_1 f_2\, d\psi\right)^2 = \tfrac{1}{2}\int_a^b\int_a^b \begin{vmatrix} f_1(x_1) & f_1(x_2) \\ f_2(x_1) & f_2(x_2) \end{vmatrix} d\psi(x_1)\, d\psi(x_2) \geq 0.$$

. *Schwarts inequality*—("=" only if f_1 and f_2 are linearly dependent.

(ii) $n=2$; $f_2\equiv\phi_2\equiv 1$. Write f, ϕ in place of f_1, ϕ_1:

(36) $$\int_a^b f\phi\, d\psi \cdot \int_a^b d\psi - \int_a^b f\, d\psi \cdot \int_a^b \phi\, d\psi$$
$$= \tfrac{1}{2}\int_a^b\int_a^b \left[f(x)-f(y)\right]\left[\phi(x)-\phi(y)\right] d\psi(x)\, d\psi(y)$$
$$\int_a^b d\psi \cdot \int_a^b f\phi\, d\psi \gtrless \int_a^b f\, d\psi \cdot \int_a^b \phi\, d\psi$$

Tchebycheff inequality (derived by him for the special case $d\psi = dx$), *where* f, ϕ *are any two functions both varying monotonically in* (a, b), *either in the same sense* (*sign* > *in* (36) *or in the opposite sense* (*sign* <). *In* (34-37) *we may replace* $d\psi(x)$ *by* $p(x)dx\left[p(x)\geq 0 \ \ in(a,\ b)\right]$.[1]

(iii) $f_i(x)=x^{i-1}$, $\phi_i(x)=F(x)x^{i-1}\left[i=1,2,\cdots\cdots n\right]$:

$$\Delta_n = \begin{vmatrix} \int_a^b F d\psi & \int_a^b F x\, d\psi & \cdots\cdots & \int_a^b F x^{n-1} d\psi \\ \int_a^b F x\, d\psi & \cdots\cdots & & \int_a^b F x^{n} d\psi \\ \cdots & \cdots & \cdots & \cdots \\ \int_a^b F x^{n-1} d\psi & \cdots\cdots & & \int_a^b F x^{2n-2} d\psi \end{vmatrix}$$
$$= \tfrac{1}{n!}\underbrace{\int_a^b\cdots\cdots\int_a^b}_{(n\ times)} \prod_{i=1}^{n} F(x_i)\, d\psi(x_i) \prod_{\substack{i,j=1 \\ i<j}}^{n} (x_i - x_j)^2.$$

1. Cf. E. Fischer, Ueber den Hadamardschen Determinontensatz, Archiv für Mathematik and Physik (3), v. 13 (1908), p. 32-49, where (34) is derived for the particular case $d\psi(x)\equiv dx$

The determinant Δ_n plays an important role in the theory of orthogonal Tchebycheff polynomials (see below). Formula (37) gives an upper limit for Δ_n:

$$(38)\qquad |\Delta_n| < \frac{1}{n!}(b-a)^{n(n-1)} M^n \left[\int_a^b d\psi(x)\right]^n$$
$$\left[M = \max. \, |F(x)| \text{ in } (a,\, b)\right].$$

Applying (13) to the above formulae, we get:

$$(39)\qquad \left(\sum_{i=1}^n u_i v_i\right)^2 \leq \sum_{i=1}^n u_i^2 \sum_{i=1}^n v_i^2$$

—*Cauchy inequality* (from (34))

$$(40)\qquad n\sum_{i=1}^n a_i b_i \gtrless \sum_{i=1}^n a_i \cdot \sum_{i=1}^n b_i$$

$$(41)\qquad \frac{\sum_{i=1}^n u_i v_i}{\sum_{i=1}^n v_i} \gtrless \frac{\sum_{i=1}^n u_i w_i}{\sum_{i=1}^n w_i} \quad (v_i,\, w_i > 0)^1$$

Formulae 40, 41 follow from (36) by means of (13), with $\sigma_i = 1$ (in (40)), $\sigma_i = v_i$ (in (41)) [$i = 1, 2, \cdots n$]. The sequences $\{(a_i),\, (b_i)\}$, $\{(u_i),\, (\frac{v_i}{w_i})\}$ are assumed to be either increasing or decreasing, the sign $\gtrless$ being chosen as in (36). Thus *all these (and many similar) inequalities have the same origin-formula* (34). Applying (13) to Hölder-Minkowski inequalities[2].

$$(42)\qquad \sum_{i=1}^n |a_i\, b_i| \leq \left\{\sum_{i=1}^n |a_i|^s\right\}^{1/s} \cdot \left\{\sum_{i=1}^n |b_i|^{s/s-1}\right\}^{s-1/s} \quad (s > 1)$$

$$\left\{\sum_{i=1}^n |a_i + b_i|^s\right\}^{1/s} \leq \left\{\sum_{i=1}^n |a_i|^s\right\}^{1/s} + \left\{\sum_{i=1}^n |b_i|^s\right\}^{1/s} \quad (s > 1)$$

we get:

$$(43)\qquad \int_a^b |f\phi|\, d\psi \leq \left\{\int_a^b |f|^s d\psi\right\}^{1/s} \cdot \left\{\int_a^b |\phi|^{\frac{s}{s-1}}\, d\psi\right\}^{s-1/s} \quad (s > 1)$$

1. Cf. 1. c. p. 73. 1-b, pp. 142, 143, 146, 194.

2. F. Rietz, Ueber Systeme integrierbarer Funktionen, Mathematische Annalen, v. 69 (1911), pp. 449-497; p. 456.

$$(44)\qquad \left\{\int_a^b |f+\phi|^s d\psi\right\}^{1/s} \leq \left\{\int_a^b |f|^s d\psi\right\}^{1/s} + \left\{\int_a^b |\phi|^s d\psi\right\}^{1/s} (s \geq 1)\ .$$

Formula (43), with $\phi \equiv 1,\ s = {}^{s_2}/_{s_1} > 1$ and f replaced by $|f|^{s_1}$, yields:

$$(45)\qquad \left\{\int_a^b |f|^{s_1} d\psi\right\}^{1/s_1} \leq \left\{\int_a^b |f|^{s_2} d\psi\right\}^{1/s_2} \cdot \left\{\int_a^b d\psi\right\}^{\frac{s_2-s_1}{s_1 s_2}}$$

$$(s_2 > s_1 > 0)\ .$$

The applications of the above inequalities to the theory of probability and mathematical statistics are many. A few illustrations follow:

(i) Consider even moments $u_{2s} = \int_{-a}^{a} x^{2s} f(x)dx = 2\int_0^a x^{2s} f(x)dx$ of a continuous unimodal symmet·· distribution over a finite interval $(-a\ ,\ a\ )$. Here (36) gives (with s_1 $<$, $d\psi(x)=f(x)dx$, $2\int_0^a f(x)dx=1$).

$$(46)\qquad u_{2s} = \int_{-a}^{a} x^{2s} f(x)dx < \frac{a^{2s}}{2s+1}$$

$$\left[s = 1,\ 2\ ,\cdots\cdots;\ f(x) \equiv f(-x)\right].$$

(ii) If ξ denotes an arbitrary constant, take in (42) $f(x) = x - \xi$, $\psi(x) =$ *law of distribution of* x *over* $(a\ ,\ b\ )$, so that $\int_a^b d\psi(x) = 1$. We get:

$$(47)\qquad v_{s_1}^{1/s_1} \leq v_{s_2}^{1/s_2} \quad \text{for } s_1 < s_2$$

$$\left(v_s^{1/s} = \left[\int_a^b |x-\xi|^s\, d\psi\right]^{1/s}\right).$$

Hence, in any distribution over any interval the quantity $v_s = \left[\int_a^b |x-\xi|^s\ d\psi\ (x)\right]^{1/s}$ *increases with* s *for any constant* ξ and, in particular,

$$u_{2s}^{1/2s} = \left[\int_a^b x^{2s}\ d\psi\right]^{1/2s} \qquad \text{also if } a \geq 0\ ,$$

$$u_s^{1/s} = \left[\int_a^b f\, d\psi\right]^{1/s}.$$ [1]

1. Paul Levy, Calcul des probabilités (Paris 1925), p. 157-58.

(iii) Apply (36) to the functions $f(x)$, $\phi(x)$ *both monotonic in* (a, b), $\psi(x)$ the same as in (ii):

(48) $E(f\phi) \gtrless E(f)E(\phi)$ (for the choice of $\gtrless$ see (36))[1].

The same formula (36) gives for *any function* $f(x)$[2]

(49) $E(f^n) > \{E(f)\}^n \quad (n = 2, 3, \cdots)$[2]

Formula (45) gives with the same $\psi(x)$:

(50) $\{E(|f|^{s_1})\}^{1/s_1} \leq \{E(|f|^{s_2})\}^{1/s_2} \quad (s_1 < s_2)$.

V. *Application of Stieltjes Integrals to Some Minimum-Problems. Given a number* $m \geq 1$ *,* M *finite points* $x_1 \leq x_2 \leq \cdots x_M$*,* M *positive quantities* $\sigma_1, \sigma_2, \cdots \sigma_M$*, and a function* $f(x)$ *with well determined values* $f(x_i)(i = 1, 2, \cdots M)$*. Find a polynomial* $P_n(x)$*, of degree not exceeding* $n(\leq M-2)$*, minimizing the expression* $\sum_{i=1}^{M} \sigma_i |f(x_i) - P_n(x_i)|^m$ *. Discuss the behavior of* $P_n(x)$ *for* $m \to \infty$. We introduce a finite interval (a, b), containing in its interior all points x_i and a monotonic non-decreasing step-wise function $\psi(x)$ with the above properties (saltus σ_i at $x = x_i$, etc.; see p. 75). Then our problem can be formulated as follows: *Find a polynomial* $P_n(x)$ *of degree not exceeding* n*, minimizing the integral*

$$\int_a^b |f(x) - P_n(x)|^m \, d\psi(x) \quad [m \geq 1] .$$

Here the advantage of Stieltjes integrals is clearly evident, for the latter problem has been discussed by G. Polya[3], D. Jackson[3] and the writer[3]. We know that a solution always exists and is unique for $m > 1$. The behavior of $P_n(x)$, when either or both m and n in-

1. G. Bohlman, Formulierung und Begründung Zweier Hulfssätze der Mathematischen Statistik, Mathematische Annalen, v. 74 (1913), pp. 341-442; p. 374-75.

2. In fact, (36) holds, with sign $>$, if $f(x) - f(y)$ and $\phi(x) - \phi(y)$ have the same sign for any x, y in a, b, which, of course, is true for $\phi(x) \equiv f(x)$.

3. (a) G. Polya, Sur un algorithme toujours convergent . . ., Comptes Rendus, v. 157 (1913) p. 840-43. (b) D. Jackson, On the Convergence of certain polynomial and trigonometric approximations. Transactions of the American Mathematical Society, v. 22 (1921), p. 158-66. (c) Idem, Note on the Convergence of Weighted Trigonometric Series, Bulletin of the American Mathematical Society, v. 29 (1923), p. 259-63. (d) J. Shohat, On the Polynomial and Trigonometric Approximation, Mathematische Annalen, v. 103 (1929), p. 157-75.

crease indefinitely, has also been discussed by the above writers. It was found that, *if* $f(x)$ *be continuous in* (a, b), *then for* n *fixed and* $m \to \infty$, $P_n(x)$ *approaches uniformly in* (a, b) *the polynomial* $\Pi_n(x)$, *of degree* $\leq n$, *of the best approximation (in Tchebycheff sense*[1]) to $f(x)$, *provided*, $\psi(x)$ *has infinitely many points of increase everywhere dense in* (a, b). *Furthermore,* $\left[\int_a^b |f(x)-P_n(x)|^m d\psi(x)\right]^{1/m}_{m\to\infty}$ *the best approximation* $E_n(f) = \max. |f(x) - \Pi_n(x)|$ *for* $a \leq x \leq b$.

This result has been supplemented by the writer (in a paper which will appear elsewhere), who showed that *the above result holds if* $\psi(x)$ *has a finite number* $M(\geq n+2)$ *points of increase.* $\Pi_n(x)$ *representing here the polynomial (of degree* $\leq n$) *giving the best approximation to* $f(x)$ *on the aggregate of the said points of increase of* $\psi(x)$. The following cases are of special interest.

(a) $n = 0$, *i. e. find a constant* X_m *minimizing the sum* $\sum_{i=1}^{M} \sigma_i |f(x_i) - X_m|^m$.

Very simple considerations show that the best approximation to $\{f(x_i)\}$ $(i = 1, 2, \cdots\cdots)$ by means of a constant is $E_0(f) = \frac{1}{2}|f(x_l) - f(x_k)|$, $f(x_l)$, $f(x_k)$ *being respectively the largest and the smallest of the* $f(x_i)$, *so that* $|f(x_l) - f(x_k)|$ *is the largest possible, and the "constant of the best approximation" is* $N_0 = \frac{1}{2}[f(x_k) + f(x_l)]$. Thus here

$$\lim_{m\to\infty} X_m = \frac{f(x_k)+f(x_l)}{2}$$

(51)
$$\lim_{m\to\infty} \left\{\sum_{i=1}^{m} \sigma_i \left[f(x_i) - X_m\right]^m\right\}^{1/m}$$

$$= \left|\frac{f(x_k)-f(x_l)}{2}\right| = \max. \left|\frac{f(x_i)-f(x_j)}{2}\right| (i, j = 1, 2, \cdots n)$$

$f(x_1) < f(x_2) < \cdots\cdots < f(x_M)$ implies:

(52)
$$\lim_{m\to\infty} X_m = \frac{f(x_M)+f(x_1)}{2}$$

$$\lim_{m\to\infty} \left\{\sum_{i=1}^{m} \sigma_i |f(x_i) - X_m|^m\right\}^{1/m} = \frac{f(x_M)-f(x_1)}{2}$$

1. That is: $E_n(f) \equiv \max. |f(x) - \Pi_n(x)| \leq \max. |f(x) - G_n(x)|$ $(a \leq x \leq b)$ where $G_n(x)$ is an arbitrary polynomial of degree $\leq n$, equality implying necessarily: $G_n \equiv \Pi_n$.

and the limiting results do not depend on $x_2, x_3, \cdots x_{n-1}$ As an illustration $f(x)=x^{2k+1}$ may serve, or, more generally, $f(x)=\sum_i A_i x^{2k_i+1}$ (all $A_i > 0$; all k_i and K are positive integers or zero)[1].

(b) $M=n+2$, *n arbitrary*. Here the writer showed (the paper will appear elsewhere):

$$(53)\ \lim_{m\to\infty} P_n(x)=\Pi_n(x)$$

$$=\frac{f_{n+1}+f_{n+2}}{2}-\frac{1}{2K}\sum_{i,j=1}^{n}(-1)^{i+j}K_{i,j}\frac{f_i-f_{i+2}}{x_i-x_{i+2}}\left(x_{n+1}^j+x_{n+2}^j-2x^j\right)$$

$$(54)\ \lim_{m\to\infty}\left[\sum_{i=1}^{n}\sigma_i\left|f_i-P_n(x_i)\right|^m\right]^{1/m}=E_n(f)$$

$$=\frac{1}{2}\left|f_k-f_{k+1}-\frac{1}{K}\sum_{i,j=1}^{n}(-1)^{i+j}\frac{f_i-f_{i+2}}{x_i-x_{i+2}}\left(x_k^j-x_{k+1}^j\right)\right|$$

$$k=1,\ 2,\cdots\cdots,M$$

$$f_i\equiv f(x_i)\ \ (i=1,\ 2,\cdots M)$$

where $K, K_{i,j}$ stand respectively for the following determinant and its minors:

$$(55)\qquad K=\begin{vmatrix} \frac{x_1^2-x_3^2}{x_1-x_3} & \frac{x_1^3-x_3^3}{x_1-x_3} & \cdots\cdots & \frac{x_1^n-x_3^n}{x_1-x_3} \\ \frac{x_2^2-x_4^2}{x_2-x_4} & \cdots\cdots & \cdots\cdots & \frac{x_2^n-x_4^n}{x_2-x_4} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{x_n^2-x_{n+2}^2}{x_n-x_{n+2}} & \cdots\cdots & \cdots\cdots & \frac{x_n^n-x_{n+2}^n}{x_n-x_{n+2}} \end{vmatrix}$$

We proceed now to show the application of Stieltjes integrals to interpolation. This must be preceded by a discussion of

VI. *Orthogonal Tchebycheff Polynomials. Theorem. Any*

---

1. Cf. D. Jackson, Note on the Median of a Set of Numbers, Bulletin of the American Mathematical Society, v. 22 (1920), p. 160-64, where the above results have been obtained for the particular case $f(x)=x$.

function $\psi(x)$, *monotonic non-decreasing on* (a, b) *—finite or infinite, and having all moments* $\gamma_k=\int_a^b x^k d\psi(x)$ $(k=0, 1, \cdots)$ *with* $\gamma_0>0$ *generates a sequence of polynomials* $\{\phi_n(x)\}$ *of degree* $n=0, 1, \cdots$ *uniquely determined by the relations*[1]: $\int_a^b \phi_m \phi_n d\psi=0$. $(m\neq n\ ;\ m, n=0, 1, \cdots)$
equivalent to $\int_a^b x^k \phi(x) d\psi(x)=0$ $(k=0, 1, \cdots n-1)$ $(n=1, 2, \cdots\cdots)$.

Proof. Take $\phi_n(x)=x^n+f_{n-1}x^{n-1}+\cdots f_1x+f_0$. The above relations lead to the following set of equations:

$$
(56)\qquad
\begin{array}{l}
f_0\gamma_0+f_1\gamma_1+\cdots\cdots\cdots f_{n-1}\gamma_{n-1}+\gamma_n=0\\
f_0\gamma_1+f_1\gamma_2+\cdots\cdots\cdots f_{n-1}\gamma_n+\gamma_{n+1}=0\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
f_0\gamma_{n-1}+f_1\gamma_n+\cdots\cdots\cdots f_{n-1}\gamma_{2n-2}+\gamma_{2n-1}=0
\end{array}
$$

The determinant Δ_n of the coefficients γ_i is (see (37)):

$$
(57)\qquad \Delta_n=\frac{1}{n!}\underbrace{\int_a^b\cdots\int_a^b}_{(n\ times)}\prod_{i=1}^{n}d\psi(x_i)\prod_{i,j=1,\ i\neq j}^{n}(x_i-x_j)^2>0
$$

which proves our statement. Add to (56) the identical relation $f_0+f_1x+\cdots\cdots\cdots+f_{n-1}x^{n-1}+(x^n-\phi_n)=0$, and for $\phi_n(x)$ we obtain the following expression:

$$
(58)\qquad \phi_n(x)=\begin{vmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_n\\ \gamma_1 & \gamma_2 & \cdots & \gamma_{n+1}\\ \cdots & \cdots & \cdots & \cdots\\ \cdots & \cdots & \cdots & \cdots\\ \cdots & \cdots & \cdots & \cdots\\ \gamma_{n-1} & \gamma_n & \cdots & \gamma_{2n-1}\\ 1 & x & \cdots & x^n\end{vmatrix} : \begin{vmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_{n-1}\\ \gamma_1 & \gamma_2 & \cdots & \gamma_n\\ \cdots & \cdots & \cdots & \cdots\\ \cdots & \cdots & \cdots & \cdots\\ \cdots & \cdots & \cdots & \cdots\\ \gamma_{n-1} & \gamma_n & \cdots & \gamma_{2n-2}\end{vmatrix}
$$

Note. If $\psi(x)$ has in (a, b) only a finite number M of points of increase, then $\Delta_n=0$ *for* $n>M$, and $\phi_n(x)$ exists only for $n=0, 1, \cdots M$ (See below (65), which in this case is a rational fraction).

The following table gives the most known and important Tchebycheff polynomials.

1. We disregard constant factors.

(a,b)	$d\psi(x) =$	$\phi(x)$ - polynomial of [Constant factors disregarded] [1]
finite	dx	Legendre: $\frac{d^n}{dx^n}\left[(x-a)^n(b-x)^n\right]$
finite	$(x-a)^{\alpha-1}(b-x)^{\beta-1}dx$ $(\alpha,\beta>0)$	Jacobi: $(x-a)^{1-\alpha}(b-x)^{1-\beta}\frac{d^n}{dx^n}\left[(x-a)^{n+\alpha-1}(b-x)^{n+\beta-1}\right]$
$(0,\infty)$	$x^{\alpha-1}e^{-kx}dx$ $(\alpha, K>0)$	Laguerre: $x^{1-\alpha}e^{kx}\frac{d^n}{dx^n}\left[x^{n+\alpha-1}e^{-kx}\right]$
$(-\infty,\infty)$	$e^{-kx^2}dx\ (K>0)$	Laplace-Hermite: $e^{kx^2}\frac{d^n}{dx^n}(e^{-kx^2})$

The polynomials $\phi_n(x)$ can be ***normalized,*** by multiplying by constant factors $a_n = 1 : \int_a^b \phi_n^2(x)d\psi$, so as to obtain ***an orthogonal and normal*** system of Tchebycheff polynomials $\{\phi_n(x) = a_n x^n + \cdots\}$ $(n = 0, 1, \cdots\cdots;\ a_n > 0)$

$$(59) \qquad \int_a^b \phi_m(x)\phi_n(x)d\psi = 0\,(m \neq n),\ = 1\,(m = n) \qquad (m, n = 0, 1, 2 \cdots\cdots)$$

The following are some of the most important properties of $\phi_n(x)$.

(a) The roots of $\phi_n(x)$ are real, distinct and lie between a, b.

(b) If all integrals $\int_a^b x^n f(x)d\psi(x)$ exist $(n = 0, 1, \cdots)$, then, by (59), we have the formal development:

$$(60) \qquad f(x) \sim \sum_{i=0}^{\infty} A_i\,\phi_i(x), \qquad \left[A_i = \int_a^b f(x)\phi_i(x)d\psi(x)\right]^{2}$$

which, regardless of its convergence or divergence, has the following remarkable property: *any "section" ("Abschnitt") of (60), i. e. the polynomial* $P_n(x) = \sum_{i=0}^{n} A_i\phi_i(x)$, *obtained by taking its first* $n+1$ *terms* $(n = 0, 1, \cdots)$ *gives the best approximation to* $f(x)$ *in* (a, b), *in the sense of least squares, i. e. it minimizes the integral* $\int_a^b [f(x) - P_n(x)]^2\,d\psi(x)$. *Moreover*

1. Cf. W. Romanowsky, Sur quelques classes nouvelles des polynomes orthogonaux, Comptes Rendus, v. 188 (1929), p. 1023-25, where new polynomials are discussed arising from Pearson's frequency curves of type IV, V, VI.

2. In the development $f(x) \sim \sum_{i=0}^{\infty} A_i\phi_i(x)$, where the $\phi_i(x)$ are not normalized, $A_i = \int_a^b f\phi_i d\psi : \int_a^b \phi_i^2 d\psi$.

$$(61)\qquad \int_a^b [f-P_n]^2 d\psi = \min. \int_a^b [f(x)-G_n(x)]^2 d\psi(x)$$
$$= \int_a^b f^2 d\psi - \sum_{i=0}^{n} A_i^2 ,$$

$G_n(x) = \sum_{i=0}^{n} g_i x^i$ denoting hereafter an *arbitrary* polynomial of degree $\leq n$. The proof is very simple. Write $G_n(x)$ as $\sum_0^n H_i \phi_i(x)$ with constant coefficient H_i, substitute this expression into $I = \int_a^b [f-G_n]^2 d\psi$, and write down the conditions of minima: $\frac{1}{2}\frac{\partial I}{\partial H_i} = 0$, which, by (59), lead to

$$H_i = \int_a^b f \phi_i \, d\psi = A_i \qquad (i = 0, 1, 2, \cdots n).$$

These coefficients A_i can be written down as linear combinations of the moments

$$(62)\qquad m_k = \int_a^b f(x) x^k d\psi(x) \quad (k = 0, 1, \cdots\cdots).$$

Introduce the symbol

$$(63)\qquad \omega(G_n) = \sum_{i=0}^{n} m_i g_i$$
$$\left(G_n(x) = \sum_{i=0}^{n} g_i x^n;\ n = 0, 1, \cdots;\ g_i \text{ arbitrary}\right)$$

Then evidently,

$$(64)\qquad A_n = \int_a^b f \phi_n \, d\psi = \omega(\phi_n)$$
$$f(x) \sim \sum_{n=0}^{\infty} \omega(\phi_n)\, \phi_n(x) ;$$

in other words, we have the following simple rule: *In the expression of $\phi_n(x)$ replace each power x^k by the corresponding moment m_k given in (62)($K = 0, 1, \cdots n$), and we obtain the coefficient A_n in (60) ($n = 0, 1, \cdots\cdot$).*

(c) $\phi_n(x)$ are denominators of the successive convergents to the continued fraction

$$(65)\qquad \int_a^b \frac{d\psi(y)}{x-y} = \frac{\lambda_1 |}{|x-c_1} - \frac{\lambda_2 |}{|x-c_2} - \cdots$$

$$(\lambda_i\ (>0),\quad c_i - const.).$$

Historically, it was the aforesaid minimum property which has lead Tchebycheff to the discovery and investigation of the *general class or orthogonal polynomials corresponding to any monotonic non-decreasing function,* while before, only isolated special cases of such polynomials have been known (polynomials of Legendre, Jacobi, Laguerre, Laplace, Hermite). Tchebycheff found these polynomials in connection with

VII. *Least-squares Interpolation.* The problem can be formulated with Tchebycheff[1] as follows: *Given the values of a certain function $y=F(x)$ at $n+1$ real, distinct points $x_1, x_2, \cdots x_{n+1}$, with the corresponding weights σ_i* [2]*. Find its value at $x=X$, assuming for y the representation $a + bx + cx^2 + \cdots + hx^m$, $(m \le n)$ so that the errors of $F(x_i)\ [i = 1, 2, \cdots n+1]$ shall have the least possible influence on the required value $F(x)$.*

Using Stieltjes integrals (which greatly simplifies Tchebycheff's analysis), we are lead to the following solution:

$$(66)\qquad \begin{gathered} F(X) = P_m(X) = \sum_{k=0}^{m} A_k\, \phi_k(X) \\ \left[A_k = \int_a^b F(x)\, \phi_k(x)\, d\psi(x) = \sum_{i=1}^{n+1} \sigma_i\, F(x_i)\, \phi_k(x_i) \right], \end{gathered}$$

where $\psi(x)$ *is the stepwise function having at $x=x_i$ a saltus $\sigma_i (i=1, 2, \cdots n+1)$, (a, b) contains in its interior all points x_i, $\{\phi_n(x)\}$ are orthogonal and normal polynomials determined by* (59), *or, which is the same, denominators of the successive convergents to the continued fraction* (65) (we disregard constant factors), *which here reduces to*

$$(67)\qquad \sum_{i=1}^{n+1} \frac{\sigma_i}{x-x_i} = \frac{\lambda_1 |}{|x-c_1} - \frac{\lambda_2 |}{|x-c_2} - \cdots\cdots$$

---

1. Tchebycheff, (a) Sur les fractions continues, Journal des Mathématiques, (2), v. III (1858), p. 289-323; (b) On the least-squares interpolation, Collected Papers, v. I, p. 473-98; (c) On interpolation with equidistant ordinates, ibid., v. II, p. 219-42 (b, c, in Russian).

2. σ_i is inversely proportional to the mean-square error of $F(x_i)$.

We see that (66) *is nothing but the first* $m+1$ *terms of the development* (60). Hence, *Tchebecheff's solution* (66) *yields the minimum of* $\int_a^b [F(x)-P_n(x)]^2 \, d\psi(x) = \sum_{i=1}^{n} \sigma_i [F(x_i)-P_n(x_i)]^2$. Moreover, for the mean-square error of (66), we get, by (59):

$$
\begin{aligned}
R^2 &= \int_a^b F^2 d\psi - \sum_{k=0}^{m} A_k^2 \\
&= \sum_{i=1}^{n+1} \sigma_i F^2(x_i) - \sum_{k=0}^{m} \left\{ \sum_{i=1}^{n+1} \sigma_i F(x_i) \phi_k(x_i) \right\}^2 .
\end{aligned}
\tag{68}
$$

The name "least-squares interpolation" is thus fully justified, and we see the *complete identity between the two problems: least-squares interpolation and approximate representation of functions by series of Tchebycheff polynomials. Whether the data are discrete and in a finite number, or the form a continuous set, the underlying principles and the resulting formulae are identical, provided we use Stieltjes integrals.* There is no need to treat the two cases separately (as one finds even in recent books on this subject) and to introduce special symbols in the first case. Another very important feature of the above solution has been indicated by Tchebycheff: If we add one more term to the expression $a+bx+\cdots+hx^m$ assumed for $y=F(x)$, we need only add one more term to $P_m(x)$ above, *without changing the preceding ones* (compare with Lagrange interpolation formula!) Formula (68) enables one to find the number of terms necessary to attain a prescribed accuracy.

Consider two special cases.

(a) *The ordinates are equidistant* $: x_{i+1}-x_i=h \; (i=1,2,\cdots n)$ *and all weights* σ_i *are equal* $(=1)$. Here Tchebycheff (1-c. 1-b. p. 91) gives very simple expressions for the polynomials $\phi_k(x)$, as well as for the coefficients A_k of (66):

$$
\begin{aligned}
\phi(z) = \Delta^k \Big[& \left(z+\frac{n-1}{2}\right)\left(z+\frac{n-3}{2}\right)\cdots\left(z+\frac{n-2k+1}{2}\right)\left(z-\frac{n+1}{2}\right)\left(z-\frac{n+3}{2}\right) \\
& \cdots \left(z-\frac{n+2k-1}{2}\right)\Big] \quad k=0,1,2,\cdots; \; z=\frac{x-\frac{1}{2}(x_1+x_n)}{x_2-x_1}; \\
& \Delta^k - k^{th} \text{ difference.}
\end{aligned}
\tag{69}
$$

$$
\begin{aligned}
u(z) \equiv F(x) = {} & \frac{1}{n}\sum_{i=1}^{n} u_i \phi_0(z) + \frac{3}{n(n^2-1^2)} \sum_{i=1}^{n} \frac{i}{1}\cdot\frac{n-i}{1} \Delta u_i \, \phi_1(z) \\
& + \frac{5}{n(n^2-1^2)(n^2-2^2)} \sum_{i=1}^{n} \frac{i(i+1)}{1\cdot 2}\cdot\frac{(n-i)(n-i-1)}{1\cdot 2} \Delta^2 u_i \, \phi_2(z) + \cdots \\
& [(u_i \equiv F(x_i)]
\end{aligned}
\tag{70}
$$

(We have replaced $n+1$ in our above formulae by n). All $\phi_n(z)$ can be easily computed by means of the relations:

$$\phi_0(z) = \Delta^0 1 = 1 \quad ; \quad \phi_1(z) = 2z \, ,$$

(71) $$\phi_k(z) = 2(2k-1)\phi_{k-1}(z) - (k-1)^2 \left[n^2 - (k-1)^2\right] \phi_{k-2}(z)$$

$$(k \geq 2)$$

(b) $m = 1$, x_i *arbitrary* $(i = 1, 2, \cdots n)$. We take in (67)

(72) $$\lambda_1 = \int_a^b d\psi(x) = \sum_{i=1}^n \sigma_i \, .$$

We get now (by successive division, for ex.)

(73) $$c_1 = \frac{\int_a^b x \, d\psi}{\int_a^b d\psi} = \frac{\gamma_1}{\gamma_0}$$

$$\left(\gamma_k = \int_a^b x^k d\psi(x) = \sum_{i=1}^n \sigma_i x_i^k \right)$$

$$\phi_0(x) = \frac{1}{\sqrt{\gamma_0}}$$

(74) $$\phi_1(x) = \frac{x_1 - c_1}{\sqrt{\int_a^b (x-c_1)^2 \, d\psi}} = \frac{\gamma_0 x - \gamma_1}{\sqrt{\gamma_0^2 \gamma_2 - \gamma_1^2}}$$

(75) $$P_1(x) = A_0 \phi_0(x) + A_1 \phi_1(x)$$

$$= \sum_{i=1}^n \frac{\sigma_i y_i}{\gamma_0} + \sum_{i=1}^n \frac{\sigma_i y_i (\gamma_0 x_i - \gamma_1)}{\gamma_0^2 \gamma_2 - \gamma_1^2} (\gamma_0 x - \gamma_1)$$

$$[\, y_i \equiv F(x_i) \,]$$

$$R^2 \text{ (mean-square error)} = \sum_{i=1}^n \sigma_i \left[F(x_i) - P_n(x_i)\right]^2$$

(76) $$= \left(\sum_{i=1}^n \frac{\sigma_i y_i}{\sqrt{\gamma_0}} \right)^2 + \left(\sum_{i=1}^n \frac{\sigma_i y_i (\gamma_0 x_i - \gamma_1)}{\sqrt{\gamma_0^2 \gamma_2 - \gamma_1^2}} \right)^2$$

(See 68)

Let $\psi(x)$ *represent a law of distribution.* Then, $\gamma_0 = 1$, $\sqrt{\gamma_0^2 \gamma_2 - \gamma_1^2} =$ standard deviation σ and the above formulae become:

$$\phi_0(x)=1\ , \qquad \phi_1(x)=\frac{x}{\sigma} \tag{77}$$

$$P_1(x)=\frac{x\int_a^b F\,x\,d\psi}{\int_a^b x^2 d\psi} \tag{78}$$

$$=\frac{x\sum_{i=1}^{n}\sigma_i\,x_i\,y_i}{\sum_{i=1}^{n}\sigma_i\,x_i^2}\qquad\left[(y_i)=F(x_i)\right]$$

$$R^2=\int_a^b F^2 d\psi-\left(\int_a^b F\phi_1\,d\psi\right)^2 \tag{79}$$

$$=\sum_{i=1}^{n}\sigma_i\,y_i^2-\frac{\left(\sum_{i=1}^{n}\sigma_i\,x_i\,y_i\right)^2}{\sigma^2}$$

One recognizes in (78, 79) formulae quite similar to those for the line of regression of y on x and for the standard error of estimate of y. Introduce

$$\sigma_x^2=\int_a^b x^2 d\psi\ ;\quad \sigma_y^2=\int_a^b y^2 d\psi,$$

$$r=\frac{\int_a^b x\,y\,d\psi}{\sqrt{\int_a^b x\,d\psi\cdot\int_a^b y\,d\psi}} \tag{80}$$

and our formulae become the classical ones:

$$P_1(x)=r\,\frac{\sigma_y}{\sigma_x}\,x\ ;\quad R=\sigma_y\,(1-r^2)^{1/2}. \tag{81}$$

We thus obtained, using Stieltjes integrals, elegant, simple and easily memorizable formulae for σ_x, σ_y and for the coefficient of correlation r. Moreover, we see by inspection (Schwartz inequality) that $-1\leq r\leq 1$, equality attainable if and only if x and y are linearly dependent. We see also that *the theory of linear regression is but a very special case of the general theory — due to Tchebecheff — of least-squares interpolation.*

1. Cf. D. Jackson. The Elementary Geometry of Function Space, American Mathematical Monthly, v. 31 (1924), p. 461-71.

Paris (France).

SIMULTANEOUS TREATMENT OF DISCRETE AND CONTINUOUS PROBABILITY BY USE OF STIELTJES INTEGRALS

By

William Dowell Baten

The object of this paper is to present several theorems pertaining to the probability that certain functions lie within certain intervals. The first theorem is a generalization of Markoff's Lemma, which is proven for the discrete and continuous cases by use of the accumulative frequency function and Stieltjes integrals. Tchebycheff's Theorem is obtained as a corollary to a very general theorem, the proof of which is based upon the first theorem. Other corollaries are given.

Three theorems, due to Guldberg, which follow are concerned with the probability that a non-negative chance variable be less than certain functions of the expected value of the variable. These are proved for the discrete and continuous cases by employing accumulative frequency functions and Stieltjes integrals. This is the first time, as far as the writer knows, the discrete and continuous cases for these theorems have been included in a single proof.

Theorem 1. If A denotes the expected value of the non-negative variable x and t is any number greater than 1, then the probability that $x \leq At^2$ is greater than $1-1/t^2$.

Proof: If x is a discrete variable with values at x_i, ($i = 1, 2, \cdots, n$) with corresponding probabilities p_i, then it is understood that the probability that x takes other values is zero. If x is a continuous variable having a probability function defined over the interval (a, b), then it is understood that the probability that x lies outside of (a, b) is zero in case (a, b) is different from $(-\infty, +\infty)$. In both cases x is a continuous variable in the interval $(-\infty, +\infty)$. Let the probability that x lies in the interval $(-\infty, x)$ be $F(x)$, with $F(-\infty) = 0$ and $F(+\infty) = 1$. Then the probability that x lies in the interval (x_1, x_2) is

$$F(x_2)-F(x_1)+\frac{1}{2}\left\{F(x_2+0)-F(x_2-0)\right\}+\frac{1}{2}\left\{F(x_1+0)-F(x_1-0)\right\}$$

where the last two limits are different from zero when there is probability different from zero at x_1 and x_2. This exists since $F(x)$ is a non-decreasing function over the interval $(-\infty, +\infty)$. In the special case when $F(x) = \int_{-\infty}^{x} f(x)dx$ where $f(x)$ is summable, $f(x)dx$ represents the probability that x lies in the interval $(x, \quad x+dx)$.

In either case, by definition

$$A = \int_{-\infty}^{\infty} x \cdot dF(x)$$

$x > At^2$ in the interval $(At^2 + \epsilon, \infty)$, where ϵ approaches 0, hence

$$A > \int_{At^2+0}^{\infty} x \cdot dF(x)$$

But

$$\int_{At^2+0}^{\infty} x dF(x) = \lim_{\epsilon \to 0} \int_{At^2+\epsilon}^{\infty} x dF(x) = \lim_{\epsilon \to 0} z_\epsilon \int_{At^2+\epsilon}^{\infty} dF(x) = \lim_{\epsilon \to 0} Z_\epsilon \cdot \lim \int_{At^2+\epsilon}^{\infty} dF(x)$$

by the first theorem of the mean, which holds for Stieltjes integrals in this case. Here $z > At^2 + \epsilon$, hence $\lim\limits_{\epsilon \to 0} z_\epsilon \geq At^2$, therefore $A > At^2 \int_{At^2+0}^{\infty} dF(x)$. But $\int_{At^2+0}^{\infty} dF(x)$ is the probability P that x is greater that At^2, hence

$$A > At^2 P, \quad Q > 1 - 1/t^2$$

where Q is the probability that $x \leq At^2$.

This theorem is a generalization of Markoff's Lemma[1], which he proved for the discrete case. The above proof takes care of the discrete case, the continuous case and the case which is a combination of the discrete and continuous.

Theorem 2. If $f(x_1, x_2, \cdots x_n)$ is a function of n independent variables, then the probability that

$$|f-k| \leq t\sqrt{E(f^2) - 2kE(f) + k^2}$$

1. "Wahrscheinlichkeitsrechnung," by Markoff. 1912. Page 54.

is greater than $1-1/t^2$; where E represents the expected value, k is a constant and $t>1$.

Proof: Let

$$y=\{f(x_1, x_2, \cdots x_n)-k\}^2 \quad \text{then}$$
$$E(y)=E(f^2)-2kE(f)+k^2$$

By theorem 1 the probability that

$$|f-k|\le t\sqrt{E(f^2)-2kE(f)+k^2}$$

is greater than $1-1/t^2$.

Corollary: If $f(x_1, x_2, \cdots x_n)=x_1+x_2+\cdots x_n$ and $k=\sum_{i=1}^{n} E(x_i)$, then theorem 2 becomes the famous Tchebycheff theorem[1]. This theorem is: If $x_1, x_2, \cdots x_n$ be n independent variables, then the probability that

$$\left|\sum_{i=1}^{n} x_i-\sum_{i=1}^{n} E(x_i)\right|\le t\sqrt{\sum_{i=1}^{n} E(x_i^2)-2\sum_{\substack{i,j=1\\ i\ne j}}^{n} E(x_i x_j)+\Big\{\sum_{i=1}^{n} E(x_i)\Big\}^2},$$

is greater than $1-1/t^2$.

This proof is by far simpler than that given by Tchebycheff, while it is similar to that given by Markoff.

In the corollary if $k=\sum_{i=1}^{n} E(x_i)$, $E(x_i)=a$, $E(x_i^2)=A$, then the probability that

$$\left|\frac{\sum x_i}{n}-a\right|\le\frac{t}{\sqrt{n}}\sqrt{A-a}$$

is greater than $1-1/t^2$.

If $f(x_1, x_2, \cdots\cdots x_n)=\sum_{i=1}^{n} x_i^s$ then the probability that

$$\left|\sum_{i=1}^{n} x_i^s-k\right|\le t\sqrt{\sum_{i=1}^{n} A_{i,2s}+2\sum_{i=1}^{n} a_{i,s}\,a_{j,s}-2k\sum_{i=1}^{n} a_{i,s}+k^2}$$

is greater than $1-1/t^2$, where the variables are independent, $E(x_i^{2s})=A_{i,2s}$; $E(x_i^s)=a_{i,s}$. If s is negative it is under-

1. "Des Valeurs Moyennes," by Tchebecheff, Journal de Math. 1867 (2). Vol. 12.

stood that x can not take on the value zero, if $s = a/b$, where a is odd and b is even, it is understood that x is non-negative.

Other interesting results may be obtained from this theorem if $f(x_1, x_2, \cdots\cdots\cdot x_n)$ represents various functions of the n independent variables and k be given different values. If f is the sum of the variables, theorem 2 is a more general theorem than Tchebycheff's theorem because of the constant k which may have values other than $\sum_{i=1}^{n} E(x_i)$.

Let x_i be the result of an individual throw of a coin, $x_i = 1$ if a head is thrown and $x_i = 0$ if a tail is thrown; then $E(x_i^2) = p \cdot 1 + q \cdot 0$, where p is the probability of a head and q is the probability of a tail. Let m represent the number of heads thrown in n throws and let $k = np \pm \sqrt{n - npq}$, then the probability that

$$\left| m - (np \pm \sqrt{n - npq}) \right| \leq t\sqrt{n}, \text{ or that}$$

$$\left| \frac{m}{n} - \left(p \pm \sqrt{\frac{1}{n} - \frac{pq}{n}}\right) \right| \leq \frac{t}{\sqrt{n}} \quad \text{, is greater than } 1 - 1/t^2 .$$

Let $t = \sqrt[4]{n}$, then the probability that

$$-\frac{1}{\sqrt[4]{n}} \pm \sqrt{\frac{1}{n} - \frac{pq}{n}} \leq \frac{m}{n} - p \leq \frac{1}{\sqrt[4]{n}} \pm \sqrt{\frac{1}{n} - \frac{pq}{n}}$$

is greater than $1 - 1/t^2$ or $1 - 1/\sqrt{n}$, which approaches unity as n increases. It is near unity for large values of n . This shows that the empirical probability approaches the true probability p as the number of throws increases, and the advantage of k .

Theorem 3: Let $u'_{n:x}$ be the expected value of the non-negative variable x raised to the power n and t any number greater than 1, then the probability that $x \leq t\sqrt[n]{u'_{n:x}}$ is greater than $1 - 1/t^n$.

Proof: Let $c > \sqrt[n]{u'_{n:x}}$, and let $F(x)$ be the probability that x lies in the interval $(-\infty, x)$, then by definition

$$u'_{n:x} = \int_{-\infty}^{\infty} x^n dF(x); \text{ and } \frac{u'_{n:x}}{c} = \int_{-\infty}^{\infty} x^n dF(x)/c^n$$

Now

$$\frac{u'_{n:x}}{c^n} > \int_{-\infty}^{\infty} x^n dF(x)/c^n = \lim_{\epsilon \to 0} \int_{c+\epsilon}^{\infty} x^n dF(x)/c^n$$

$$= \lim_{\epsilon \to 0} \frac{(z_\epsilon)^n}{c^n} \cdot \lim_{\epsilon \to 0} \int_{c+\epsilon}^{\infty} dF(x) > 1 \cdot \int_{c+0}^{\infty} dF(x),$$

by the first theorem of the mean, and since $\lim\limits_{\epsilon \to 0} \frac{(z_\epsilon)^n}{c^n} \geq 1$.

Since $\int_{c+0}^{\infty} dF(x)$ is the probability P that x is greater than c,

$$\frac{u'_{n:x}}{c^n} > p \quad \text{or} \quad Q > 1 - \frac{u'_{n:x}}{c^n}$$

where Q is the probability that $x \leq c$.

Let $t = \frac{c}{\sqrt[n]{u'_{n:x}}}$, then $\frac{1}{t^n} = \frac{u'_{n:x}}{c^n}$, hence

$$Q > 1 - 1/t^n .$$

But Q becomes the probability that $x \leq \sqrt[n]{u'_{n:x}}$, since c was any number greater than $\sqrt[n]{u'_{n:x}}$.

Let $y = |x - k|$, then theorem 3 becomes: If $u'_{n:y}$ is the expected value of $|x - k|^n$ and t is greater than 1, then the probability that $|x - k|$ does not surpass the multiple $t\sqrt[n]{u'_{n:y}}$, is greater than $1 - 1/t^n$, where k is a constant.

If $k = \int_{-\infty}^{\infty} x \cdot dF(x)$, then $u'_{n:y}$ becomes $u'_{n:\bar{x}}$ and theorem 3 states that the difference $|x - k|$ does not surpass the multiple $t\sqrt[n]{u'_{n:\bar{x}}}$, is greater than $1 - 1/t^n$. In this special case theorem 3 becomes Guldberg's theorem[1], but this is more general than his theorem, for it includes the continuous case, the discrete case and the case which is a combination of the discrete and continuous.

If $y = |f(x) - k|$ is used for the variable, a more general theorem is obtained. Here $f(x)$ is a function of x. Of course, the probability law for $f(x)$ must be secure from that of x if the continuous case is under consideration. Certain restrictions must be placed upon $f(x)$ concerning continuity, summability and concerning the inverse.

---

1. "Sur un théorèmede M. Markoff," by Alf. Guldberg. Compte Rendue, Vol. 175. (1922) page 679.

Theorem 4. The probability that the difference $|x - m|$ is not greater than the multiple $t\,u_{r:x}$, $t > 1$, is greater than $1 - \left(\sqrt[n]{u'_{n:x}}/u'_{r:x}\right)^n \left(1/t^n\right)$, where $u_{r:x}$ is the expected value of $|x-m|^r$, and m is the expected value of x.

Theorem 5. The probability that the positive quantity x does not surpass the multiple tm, $(t > 1)$, is greater than $1-\left(\frac{\sqrt[n]{u'_{n:x}}}{m}\right)^n \cdot \frac{1}{(t-1)^n}$,

where $u_{n:x}$ is the expected value of $|x-m|^n$ and $m = E(x)$.

These last two theorems are due to Guldberg[1] for the discrete case. By the method used in theorem 3 these can be proven for the continuous case, the discrete case, and the case which is a combination of the discrete and continuous.

---

1. "Sur quelques inéqualités des le calcul de probabilités," by Guldberg. Comp. Rend. Vol. 175 (1922), p. 1382.
   "Sur le théorème de, Tchebecheff," by Guldberg. Comptes Rendue, Vol. 175 (1922), p. 418.

EDITORIAL

FUNDAMENTALS OF THE THEORY OF SAMPLING

I. Sampling from a Limited Supply

We shall consider first a population of s individuals, in which each individual possesses a common attribute that can be measured quantitatively. The sum of the associated variates may be expressed as follows:

$$x_1 + x_2 + x_3 + \cdots x_s = \sum^{s} x = sM_x$$

From this so-called *parent population* it is possible to select $\binom{s}{r}$ different *samples*, each consisting of r individuals, ($r \leq s$). These samples may be ordered after any fashion, and the algebraic sum of the variates for the respective samples may be designated

$$\begin{array}{lllllll}
z_1 = & x_1 & +x_2 & +x_3 & \cdots & +x_r = & \sum^{r:1} x \\
z_2 = & x_2 & +x_3 & +x_4 & \cdots & +x_{r+1} = & \sum^{r:2} x \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \\
z_{\binom{s}{r}} = & x_{s-r+1} & +x_{s-r+2} & & \cdots & +x_s = & \sum^{r:\binom{s}{r}} x
\end{array}$$

Thus, while $\sum^{s} x$ represents the sum of all the s variates in the parent population, $\sum^{r:i} x$ designates the sum of the r variates occurring in the i th sample.

We face now the problem of describing adequately, from a statistical point of view, the distribution of these $\binom{s}{r}$ values of z , that is to say, we must express the moments $\mu_{n:z}$ in terms of the moments of the parent population, $\mu_{n:x}$.

By definition $\qquad M_z = \dfrac{\sum z}{\binom{s}{r}}$

Since each value of z will contribute r terms to the value of $\sum z$, this latter expression will consist of $r\cdot\binom{s}{r}$ terms involving each of the s variates of the parent population alike. Therefore, each variate, x_i ($i = 1, 2, 3, \ \ldots \ s$), will occur in the expression for $\sum z$ exactly $\frac{r}{s}\cdot\binom{s}{r}$ times. Consequently

(1) $$M_z = \frac{\sum z}{\binom{s}{r}} = \frac{1}{\binom{s}{r}}\cdot\frac{r}{s}\cdot\binom{s}{r}\left\{x_1 + x_2 + \cdots x_s\right\} = \frac{r}{s}\sum^{s} x = rM_x$$

We shall now investigate the values of

$$\mu_{n:z} = \frac{\sum \bar{z}^{\,n}}{\binom{s}{r}}$$

where we choose to represent a deviation from the mean as

$$\bar{z}_i = z_i - M_z$$

Observing that

$$\bar{z}_1 = z_1 - M_z = x_1 + x_2 + \cdots x_r - rM_x = \bar{x}_1 + \bar{x}_2 + \cdots \bar{x}_r$$

we note that

$$\bar{z}_1^2 = \sum^{r:1} \bar{x}^2 + 2\sum^{r:1} \bar{x}_i\bar{x}_j$$

$$\bar{z}_2^2 = \sum^{r:2} \bar{x}^2 + 2\sum^{r:2} \bar{x}_i\bar{x}_j$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\bar{z}^2_{\binom{s}{r}} = \sum^{r:\binom{s}{r}} \bar{x}^2 + 2\sum^{r:\binom{s}{r}} \bar{x}_i\bar{x}_j$$

Therefore

$$\mu_{2:z} = \frac{\sum \bar{z}^2}{\binom{s}{r}} = \frac{1}{\binom{s}{r}}\left\{\frac{r\cdot\binom{s}{r}}{s}\sum^{s} \bar{x}^2 + 2\frac{\binom{r}{2}\binom{s}{r}}{\binom{s}{2}}\sum^{s} \bar{x}_i\bar{x}_j\right\},$$

or, writing

$$\rho_i = \frac{r^{(i)}}{s^{(i)}} = \frac{r(r-1)(r-2)\cdot\,\cdots\,\cdot(r-\overline{i-1})}{s(s-1)(s-2)\qquad\qquad(s-\overline{i-1})},$$

(2a) $$\mu_{2:z} = 2!\left\{\rho_1\frac{\sum^{s}\bar{x}^2}{2!} + \rho_2\frac{\sum^{s}\bar{x}_i\bar{x}_j}{(1!)^2}\right\}$$

By utilizing further the multinomial theorem, it follows easily that

$$\mu_{3:x} = 3!\left\{\rho_1 \frac{\sum^s \bar{x}^3}{3!} + \rho_2 \frac{\sum^s \bar{x}_i^2 \bar{x}_j}{2!\,1!} + \rho_3 \frac{\sum^s \bar{x}_i \bar{x}_j \bar{x}_k}{(1!)^3}\right\} \tag{3a}$$

$$\mu_{4:x} = 4!\left\{\rho_1 \frac{\sum^s \bar{x}^4}{4!} + \rho_2 \frac{\sum^s \bar{x}_i^3 \bar{x}_j}{3!\,1!} + \rho_2 \frac{\sum^s \bar{x}_i^2 \bar{x}_j^2}{(2!)^2}\right. \tag{4a}$$

$$\left. + \rho_3 \frac{\sum^s \bar{x}_i^2 \bar{x}_j \bar{x}_k}{2!(1!)^2} + \rho_4 \frac{\sum^s \bar{x}_i \bar{x}_j \bar{x}_k \bar{x}_l}{(1!)^4}\right\}$$

etc.

The rule for writing down the terms is as follows: The number of terms in the expression for $\mu_{n:x}$ equals the number of partitions that can be formed from the integer n. The subscript of ρ equals the number of elements in the corresponding partition, and exponents of $\bar{x}$ and the factorials in the denominators are in fact the elements of the partitions.

Our next problem is to express the summations in terms of moments of the parent population, $\mu_{n:x}$.

First order summation

$$\sum^s \bar{x} = s\mu_{1:x} = 0$$

Second order summations

$$\sum \bar{x}^2 = s\mu_{2:x}$$

$$2\sum \bar{x}_i \bar{x}_j = -s\mu_{2:x}$$

since $$\left(\sum^s \bar{x}\right)^2 = 0 = \sum^s \bar{x}^2 + 2\sum^s \bar{x}_i \bar{x}_j$$

Third order summations

$$\sum^s \bar{x}^3 = s\mu_{3:x}$$

$$\sum^s \bar{x}_i^2 \bar{x}_j = -s\mu_{3:x}$$

$$3\sum^s \bar{x}_i \bar{x}_j \bar{x}_k = s\mu_{3:x} \;.$$

since $\sum^{s} \bar{x}^{2} \sum^{s} \bar{x} = 0 = \sum^{s} \bar{x}^{3} + \sum^{s} \bar{x}_i^{2} \bar{x}_j$

and $\left(\sum^{s} \bar{x}\right)^{3} = 0 = \sum^{s} \bar{x}^{3} + 3\sum^{s} \bar{x}_i^{2} \bar{x}_j + 6\sum^{s} \bar{x}_i \bar{x}_j \bar{x}_k$

Fourth order summations

$$\sum^{s} \bar{x}^{4} = s\mu_{4:x}$$

$$\sum^{s} \bar{x}_i^{3} \bar{x}_j = -s\mu_{4:x}$$

$$2\sum^{s} \bar{x}_i^{2} \bar{x}_j^{2} = -s\mu_{4:x} + s^{2}\mu_{2:x}^{2}$$

$$2\sum^{s} \bar{x}_i^{2} \bar{x}_j \bar{x}_k = 2s\mu_{4:x} - s^{2}\mu_{2:x}^{2}$$

$$8\sum^{s} \bar{x}_i \bar{x}_j \bar{x}_k \bar{x}_l = -2s\mu_{4:x} + s^{2}\mu_{2:x}^{2}$$

Utilizing these summations, (2a), (3a) and (4a) may be written

(2) $\mu_{2:\bar{x}} = s\mu_{2:x}\{\rho_1 - \rho_2\}$

(3) $\mu_{3:\bar{x}} = s\mu_{3:x}\{\rho_1 - 3\rho_2 + 2\rho_3\}$

(4) $\mu_{4:\bar{x}} = s\mu_{4:x}\{\rho_1 - 7\rho_2 + 12\rho_3 - 6\rho_4\} + 3s^{2}\mu_{2:x}^{2}\{\rho_2 - 2\rho_3 + \rho_4\}$.

Continuing after this fashion, one can show after a lavish use of symmetric functions that

(5) $\mu_{5:\bar{x}} = s\mu_{5:x}\{\rho_1 - 15\rho_2 + 50\rho_3 - 60\rho_4 + 24\rho_5\}$
$+ 10s^{2}\mu_{3:x}\mu_{2:x}\{\rho_2 - 4\rho_3 + 5\rho_4 - 2\rho_5\}$,

(6) $\mu_{6:\bar{x}} = s\mu_{6:x}\{\rho_1 - 31\rho_2 + 180\rho_3 - 390\rho_4 + 360\rho_5 - 120\rho_6\}$

$+ 15s^{2}\mu_{4:x}\mu_{2:x}\{\rho_2 - 8\rho_3 + 19\rho_4 - 18\rho_5 + 6\rho_6\}$

$+ 10s^{2}\mu_{3:x}^{2}\{\rho_2 - 6\rho_3 + 13\rho_4 - 12\rho_5 + 4\rho_6\}$

$+ 15s^{3}\mu_{2:x}^{3}\{\rho_3 - 3\rho_4 + 3\rho_5 - \rho_6\}$.

$$(7)\quad \mu_{7:s} = s\mu_{7:x}\{\rho_1 - 63\rho_2 + 602\rho_3 - 2100\rho_4 + 3360\rho_5 - 2520\rho_6 + 720\rho_7\}$$
$$+ 21s^2\mu_{5:x}\mu_{2:x}\{\rho_2 - 16\rho_3 + 65\rho_4 - 110\rho_5 + 84\rho_6 - 24\rho_7\}$$
$$+ 35s^2\mu_{4:x}\mu_{3:x}\{\rho_2 - 10\rho_3 + 35\rho_4 - 56\rho_5 + 42\rho_6 - 12\rho_7\}$$
$$+ 105s^3\mu_{3:x}\mu_{2:x}^2\{\rho_3 - 5\rho_4 + 9\rho_5 - 7\rho_6 + 2\rho_7\}.$$

$$(8)\quad \mu_{8:s} = s\mu_{8:x}\{\rho_1 - 127\rho_2 + 1932\rho_3 - 10206\rho_4 + 25200\rho_5 - 31920\rho_6 + 20160\rho_7 - 5040\rho_8\}$$
$$+ 28s^2\mu_{6:x}\mu_{2:x}\{\rho_2 - 32\rho_3 + 211\rho_4 - 570\rho_5 + 750\rho_6 - 480\rho_7 + 120\rho_8\}$$
$$+ 56s^2\mu_{5:x}\mu_{3:x}\{\rho_2 - 18\rho_3 + 97\rho_4 - 240\rho_5 + 304\rho_6 - 192\rho_7 + 48\rho_8\}$$
$$+ 35s^2\mu_{4:x}^2\{\rho_2 - 14\rho_3 + 73\rho_4 - 180\rho_5 + 228\rho_6 - 144\rho_7 + 36\rho_8\}$$
$$+ 210s^3\mu_{4:x}\mu_{2:x}^2\{\rho_3 - 9\rho_4 + 27\rho_5 - 37\rho_6 + 24\rho_7 - 6\rho_8\}$$
$$+ 280s^3\mu_{3:x}^2\mu_{2:x}\{\rho_3 - 7\rho_4 + 19\rho_5 - 25\rho_6 + 16\rho_7 - 4\rho_8\}$$
$$+ 105s^4\mu_{2:x}^4\{\rho_4 - 4\rho_5 + 6\rho_6 - 4\rho_7 + \rho_8\}.$$

It is convenient, at this point, to define the "*n th sampling polynomial*" as follows:

$$(9)\quad P_n(\rho) = D_x^n \log(\rho e^x + 1 - \rho)\Big]_{x=0}$$

If we place $y = \log(\rho e^{x} + 1 - \rho)$, then

$$y^{(1)} = \frac{\rho e^{x}}{\rho e^{x} + 1 - \rho} \quad \text{or} \quad (\rho e^{x} + 1 - \rho)\, y^{(1)} = \rho e^{x}$$

Taking the n th derivative of both sides by utilizing Leibnitz' Theorem, we obtain

$$(\rho e^{x} + 1 - \rho)\, y^{(n+1)} + \binom{n}{1} \rho e^{x} y^{(n)} + \binom{n}{2} \rho e^{x} y^{(n-1)} + \cdots = \rho e^{x}.$$

Placing $x = 0$ in this equation yields, by definition,

$$P_{n+1}(\rho) + \binom{n}{1} \rho P_n(\rho) + \binom{n}{2} \rho P_{n-1}(\rho) + \cdots = \rho$$

That is, for $n = 0,\ 1,\ 2,\ \cdots\cdots$

$$P_1(\rho) = \rho$$

$$P_2(\rho) + \rho P_1(\rho) = \rho$$

$$P_3(\rho) + 2\rho P_2(\rho) + \rho P_1(\rho) = \rho$$

$$P_4(\rho) + 3\rho P_3(\rho) + 3\rho P_2(\rho) + \rho P_1(\rho) = \rho$$

etc.

Thus:

$$(10) \quad \begin{cases} P_1(\rho) = \rho \\ P_2(\rho) = \rho - \rho^2 \\ P_3(\rho) = \rho - 3\rho^2 + 2\rho^3 \\ P_4(\rho) = \rho - 7\rho^2 + 12\rho^3 - 6\rho^4 \\ P_5(\rho) = \rho - 15\rho^2 + 50\rho^3 - 60\rho^4 + 24\rho^5 \\ P_6(\rho) = \rho - 3\rho^2 + 180\rho^3 - 390\rho^4 + 360\rho^5 - 120\rho^6 \\ P_7(\rho) = \rho - 63\rho^2 + 602\rho^3 - 2100\rho^4 + 3360\rho^5 - 2520\rho^6 + 720\rho^7 \\ P_8(\rho) - \rho - 127\rho^2 + 1932\rho^3 - 10206\rho^4 + 25200\rho^5 - 31920\rho^6 \\ \qquad \left[+ 20160\rho^7 - 5040\rho^8 \right] \end{cases}$$

etc.

The law of formation of the coefficients is obvious: for if $c_{i:n}$ designates the coefficient of ρ^i in the expression for $P_n(\rho)$,

$$c_{i:n} = i c_{i:n-1} - (i-1) c_{i-1:n-1}$$

Comparing the polynomials of equations (9) with formulae (2) to (8) inclusive, suggests writing the expressions for $\mu_{n:s}$ in the following symbolic form:

$$\mu_{2:s} = 2!\left\{P_2 \frac{s\mu_{2:x}}{2!}\right\}$$

$$\mu_{3:s} = 3!\left\{P_3 \frac{s\mu_{3:x}}{3!}\right\}$$

(11)
$$\mu_{4:s} = 4!\left\{P_4 \frac{s\mu_{4:x}}{4!} + \frac{P_2^2}{2!} \frac{s\mu^2_{2:x}}{(2!)^2}\right\}$$

$$\mu_{5:s} = 5!\left\{P_5 \frac{s\mu_{5:x}}{5!} + P_3 P_2 \frac{s^2\mu_{3:x}\mu_{2:x}}{3!\,2!}\right\}$$

$$\mu_{6:s} = 6!\left\{P_6 \frac{s\mu_{6:x}}{6!} + P_4 P_2 \frac{s^2\mu_{4:x}\mu_{2:x}}{4!\,2!} + \frac{P_3^2}{2!} \frac{s^2\mu^2_{3:x}}{(3!)^2} + \frac{P_2^3}{3!} \frac{s^3\mu^3_{2:x}}{(2!)^3}\right\}$$

etc.

By P_n we understand an expression derived from the sampling polynomial, $P_n(\rho)$, by writing ρ^i as ρ_i. Thus,

$$P_4(\rho) = \rho - 7\rho^2 + 12\rho^3 - 6\rho^4 \text{, whereas}$$

$$P_4 = \rho_1 - 7\rho_2 + 12\rho_3 - 6\rho_4$$

Again, since

$$P_3(\rho)\cdot P_1(\rho)\cdot P_1(\rho) = (\rho - 3\rho^2 + 2\rho^3)\rho\cdot\rho = \rho^3 - 3\rho^4 + 2\rho^5,$$

$$P_3 P_1^2 = \rho_3 - 3\rho_4 + 2\rho_5$$

The number of terms in the expression for $\mu_{n:s}$ will equal the number of partitions that can be formed from the integer n. The subscripts of the P and μ factors for any selected term correspond to the elements of the corresponding partition, and the exponent of s equals the number of elements in the partition. The factorials beneath the μ factors agree with the order of these moments, and the factorials appearing occasionally under the P factors depend upon the

number of times that any P is repeated as a factor in that term. All terms arising from a partition in which unity is an element have been neglected, since such terms will contain $\mu_{1:x}$ as a factor and consequently be equal to zero.

Illustration I. For the parent population we shall select the following (it will be noted that graphically the ordinates terminate on the hypotenuse of an isosceles right triangle):

TABLE I

Parent Population

x	f_x
1	24
2	23
3	22
4	21
5	20
..	..
22	3
23	2
24	1
Total	300

The mean, standard deviation and moments about the mean for this distribution are as follows:

$$\begin{array}{llll}
M_x = 8.666 & & & \\
\mu_{2:x} = 33.222 & & \sigma_x = & 5.76387 \\
\mu_{3:x} = 108.526 & & \alpha_{3:x} = & .566749 \\
\mu_{4:x} = 2642.27 & & \alpha_{4:x} = & 2.39398 \\
\mu_{5:x} = 20525.2 & & \alpha_{5:x} = & 5.69279 \\
\mu_{6:x} = 322570 & & \alpha_{6:x} = & 27.3878
\end{array}$$

It may well be remarked at this point that the standard variate corresponding to an observed variate, x_i is

$$(12) \qquad t_i = \frac{x_i - M_x}{\sigma_x} = \frac{\bar{x}_i}{\sigma_x} ,$$

and is consequently an *abstract number*. The n th moment of the standard variates is also without unit, i. e.

$$\alpha_{n:x} = \frac{\sum t^n}{N} = \frac{1}{N\sigma_x^n} \sum \bar{x}_i^n = \frac{\mu_{n:x}}{\sigma_x^n} \tag{13}$$

In dealing with distributions one should always bear in mind that the mean and standard deviation determine merely the *position* of the centroid vertical and the *scale* of the distribution, but that the standard moments are influenced by the *shape* of the distribution alone. Consequently a study of the mathematical representation of frequency distributions is essentially an investigation concerning the standard moments of observed and theoretical distributions.

From the above parent population it would be possible to select $\binom{300}{25}$ samples, each consisting of 25 individuals. To describe the distribution of these sampes, we proceed as follows:

$$\rho_1 = \qquad = .08333$$

$$\rho_2 = \rho_1 \cdot \frac{24}{299} = .0066 \ 8896 \ 32$$

$$\rho_3 = \rho_2 \cdot \frac{23}{298} = .0005 \ 1626 \ 226$$

$$\rho_4 = \rho_3 \cdot \frac{22}{297} = .0000 \ 3824 \ 1649$$

$$\rho_5 = \rho_4 \cdot \frac{21}{296} = .0000 \ 0271 \ 3090 \ 0$$

$$\rho_6 = \rho_5 \cdot \frac{20}{295} = .0000 \ 0018 \ 3938 \ 31$$

$P_2 = .0766 \ 4437 \ 0$	$P_6 = -.0450 \ 5692 \ 2$
$P_3 = .0642 \ 9896 \ 8$	$P_4 = .0032 \ 3772 \ 45$
$P_4 = .0424 \ 7628 \ 8$	$P_3^2 = .0040 \ 5670 \ 98$
$P_2^3 = .0056 \ 9468 \ 03$	$P_2^3 = .0004 \ 0949 \ 264$
$P_5 = .0065 \ 8261 \ 36$	
$P_3 P_2 = .0048 \ 0969 \ 62$	

$M_x = 216.66$

$\mu_{2:x} = 763.88$ $\qquad \sigma_x = 27.6385$

$\mu_{3:x} = 2093.43$ $\qquad \alpha_{3:x} = 0991550$

$\mu_{4:x} = 1730700$ $\qquad \alpha_{4:x} = 2.96594$

$\mu_{5:x} = 15647600$ $\qquad \alpha_{5:x} = .970225$

$\mu_{6:x} = 6503500000$ $\qquad \alpha_{6:x} = 14.5900$

As a check on this theory, three hundred Hollerith cards were punched with numbers corresponding to the three hundred variates of the parent population. The cards were thoroughly shuffled and then placed in a tabulating machine. After twenty-five cards had run through this electric tabulator, their total was recorded. By repeating this procedure one thousand samples were readily obtained and the results are presented below.

TABLE II

Distribution of the Totals of Samples of Twenty-five Variates Selected at Random from the Parent Population of Table I

Class	*Frequency*
120–	6
140–	28
160–	78
180–	179
200–	273
220–	229
240–	124
260–	56
280–	20
300–	7
Total	1000

In this observed distribution it is found that

$M = 215.84$ $\qquad \sigma = 30.8505$

$\alpha_3 = .1556\ 56$ $\qquad \alpha_5 = 1.39471$

$\alpha_4 = 3.18939$ $\qquad \alpha_6 = 15.8603$

The significance of the differences that exist between these functions and the values of M_2, σ_2 and $\alpha_{n,2}$ given above will be considered in a subsequent paper.

The unmodified moments, ν, for the preceding observed distribution were corrected for grouping by means of the following formula:

$$\begin{aligned} \mu_n = \nu_n - \binom{n}{2}\frac{1-\frac{1}{k^2}}{12}\nu_{n-2} + \binom{n}{4}\frac{\left(1-\frac{1}{k^2}\right)\left(7-\frac{3}{k^2}\right)}{240}\nu_{n-4} \\ - \binom{n}{6}\frac{\left(1-\frac{1}{k^2}\right)\left(31-\frac{18}{k^2}+\frac{3}{k^4}\right)}{1344}\nu_{n-6} + \cdots \end{aligned} \tag{14}$$

where k represents the number of different equidistant variates that can appear in each class. In our case, $k = 20$. Sheppard's corrections will appear as a special case of this formula by permitting k to approac infinity. Thus

$$\mu_n = \nu_n - \binom{n}{2}\frac{1}{12}\nu_{n-2} + \binom{n}{4}\frac{7}{240}\nu_{n-4} - \binom{n}{6}\frac{31}{1344}\nu_{n-6} + \cdots * \tag{15}$$

At first thought one is apt to be surprised in observing that the distribution of samples appearing in Table II is so nearly "normal," whereas the samples were taken from a right-triangular parent population. As an even more extreme case, I may mention that a group of students chose arbitrarily the following most unusual distribution for a parent population:

TABLE III

x	f_x
15	9
3	2
29	43
405	189
1710	37
Total	280

*Compare with formulae (2b), page 94, Handbook of Mathematical Statistics.

and found that the distribution of the totals of 1000 samples of twenty-five variates each was as follows:

TABLE IV

Class	Freq.
5000–	2
7000–	54
9000–	203
11000–	310
13000–	254
15000–	130
17000–	36
19000–	9
21000–	2
Total	1000

As a matter of fact, if r is fifty or greater and s is at least ten times as large as r, the parent population has relatively little control over the shape of the distribution of samples. But before investigating the limit towards which distributions of samples approach in shape, it is well to present a second illustration of the theory so far developed.

Illustration II. Pearson's Hypergeometric Series.

If from a bag containing qs black and ps white balls, r balls are withdrawn without replacements, the chances that the r balls withdrawn will contain 0, 1, 2, . . ., x, . . . r white balls are given by the successive terms of the hypergeometric series

$$(16) \quad \frac{1}{\binom{s}{r}}\left\{\binom{qn}{r}+\binom{qn}{r-1}\binom{pn}{1}+\cdots\binom{qn}{r-x}\binom{pn}{x}+\cdots\binom{pn}{r}\right\}$$

A distribution of this type is equivalent to the simplest case that can arise in accordance with the theory of sampling, that is, by assuming that each variate of the parent population is equal to either zero or one, and that p denotes the proportion of the s variates that have

unit value. The moments of the parent population are found as follows:

TABLE V

Parent Population for Hypergeometric (and Binomial) Series

x	f_x	xf_x	$x-M_x=\bar{x}$	$(x-M_x)^n f_x = \bar{x}^n f_x$
0	$(1-p)s$	0	$-p$	$(-1)^n p^n (1-p)s$
1	ps	ps	$1-p$	$p(1-p)^n \cdot s$
Total	s	ps		$p(1-p)s\{(1-p)^{n-1}+(-1)^n p^{n-1}\}$

Therefore

$$(17) \quad \mu_{n:x} = p(1-p)\{(1-p)^{n-1}+(-1)^n p^{n-1}\} = pq\{q^{n-1}+(-1)^n p^{n-1}\},$$

where $(p+q=1)$

In numerical problems this formula should be used ordinarily as it stands, although for algebraic purposes we may use frequently the forms

$$\mu_{1:x} = 0$$
$$\mu_{2:x} = pq = p(1-p)$$
$$\mu_{3:x} = pq\,(q^2-p^2) = p(1-p)(1-2p)$$
$$\mu_{4:x} = pq\,(q^3+p^3) = p(1-p)(1-3p+3p^2)$$

etc.

Using formulae 2, . . ., we may write the moments for the hypergeometric series as follows:

$$\mu_{2:t} = s\mu_{2:x}\{\rho_1-\rho_2\}$$
$$\mu_{3:t} = s\mu_{3:x}\{\rho_1-3\rho_2+2\rho_3\}$$

etc.

or if one prefers

$$\mu_{2:2} = spq \left\{ \frac{r}{s} - \frac{r^{(2)}}{s^{(2)}} \right\}$$

$$\mu_{3:2} = spq\,(q^2 - p^2) \left\{ \frac{r}{s} - 3\,\frac{r^{(2)}}{s^{(2)}} + 2\,\frac{r^{(3)}}{s^{(3)}} \right\}$$

$$\mu_{4:2} = spq\,(q^3 + p^3) \left\{ \frac{r}{s} - 7\,\frac{r^{(2)}}{s^{(2)}} + 12\,\frac{r^{(3)}}{s^{(3)}} - 6\,\frac{r^{(4)}}{s^{(4)}} \right\}$$

$$+ 3s^2p^2q^2 \left\{ \frac{r^{(2)}}{s^{(2)}} - 2\,\frac{r^{(3)}}{s^{(3)}} + \frac{r^{(4)}}{s^{(4)}} \right\}$$

etc.

These will be found equivalent to those given by Pearson*, namely

$$\mu_2 = \frac{\alpha\beta\,(s+\alpha)(s+\beta)}{s^2(s-1)}$$

$$\mu_3 = \frac{\alpha\beta(s+\alpha)(s+\beta)(s+2\alpha)(s+2\beta)}{s^3(s-1)(s-2)}$$

$$\mu_4 = \frac{m_2\,(s^2+m_1 s+m_2)}{s(s-1)(s-2)(s-3)} \Big\{ s^4 + s^3\,(3m_2 + 6m_1 + 1)$$

$$+ 3s^2(m_1 m_2 + 2m_1^2 + 2m_2)$$

$$+ 3s\,m_2\,(m_2 + 6m_1) + 10\,m_2^2 \Big\}$$

where

$$\alpha = -r \qquad \beta = -ps$$

$$m_1 = \alpha + \beta \qquad m_2 = \alpha\beta$$

II. Sampling from an Unlimited Supply

Referring to the formula of the first part of this paper, we observe that as s approaches infinity, r remaining finite,

---

*Lond., Edinburgh and Dublin Phil. Mag., Jan.-June, 1899, page 236.

$$
(18)\begin{cases}
M_z = rM_x \\
\mu_{2:z} = r\mu_{2:x} \\
\mu_{3:z} = r\mu_{3:x} \\
\mu_{4:z} = r\mu_{4:x} + 3r^{(2)}\mu_{2:x}^2 \\
\mu_{5:z} = r\mu_{5:x} + 10r^{(2)}\mu_{3:x}\mu_{2:x} \\
\mu_{6:z} = r\mu_{6:x} + 15r^{(2)}\mu_{4:x}\mu_{2:x} + 10r^{(2)}\mu_{3:x}^2 + 15r^{(3)}\mu_{2:x}^3 \\
\mu_{7:z} = r\mu_{7:x} + 21r^{(2)}\mu_{5:x}\mu_{2:x} + 35r^{(2)}\mu_{4:x}\mu_{3:x} \\
\qquad + 105r^{(3)}\mu_{3:x}\mu_{2:x}^2 \\
\mu_{8:z} = r\mu_{8:x} + 28r^{(2)}\mu_{6:x}\mu_{2:x} + 56r^{(2)}\mu_{5:x}\mu_{3:x} + 35r^{(2)}\mu_{4:x}^2 \\
\qquad + 210r^{(3)}\mu_{4:x}\mu_{2:x}^2 + 280r^{(3)}\mu_{3:x}^2\mu_{2:x} + 105r^{(4)}\mu_{2:x}^4
\end{cases}
$$

From these the following equations may be obtained:

$$
(19)\begin{cases}
\mu_{2:z} = r\mu_{2:x} \\
\mu_{3:z} = r\mu_{3:x} \\
\mu_{4:z} - 3\mu_{2:z}^2 = r\{\mu_{4:x} - 3\mu_{2:x}^2\} \\
\mu_{5:z} - 10\mu_{3:z}\mu_{2:z} = r\{\mu_{5:x} - 10\mu_{3:x}\mu_{2:x}\} \\
\mu_{6:z} - 15\mu_{4:z}\mu_{2:z} - 10\mu_{3:z}^2 + 30\mu_{2:z}^3 = r\{\mu_{6:x} - 15\mu_{4:x}\mu_{:} \\
\qquad - 10\mu_{3:x}^2 + 30\mu_{2:x}^3\} \\
\mu_{7:z} - 21\mu_{5:z}\mu_{2:z} - 35\mu_{4:z}\mu_{3:z} + 210\mu_{3:z}\mu_{2:z}^2 = r\{\mu_{7:x} \\
\qquad - 21\mu_{5:x}\mu_{2:x} - 35\mu_{4:x}\mu_{3:x} + 210\mu_{3:x}\mu_{2:x}^2\} \\
\mu_{8:z} - 28\mu_{6:z}\mu_{2:z} - 56\mu_{5:z}\mu_{3:z} - 35\mu_{4:z}^2 + 420\mu_{4:z}\mu_{2:z}^2 \\
\qquad + 560\mu_{3:z}^2\mu_{2:z} - 630\mu_{2:z}^4 = r\{\mu_{8:x} - 28\mu_{6:x}\mu_{2:x} \\
\qquad - 56\mu_{5:x}\mu_{3:x} - 35\mu_{4:x}^2 + 420\mu_{4:x}\mu_{2:x}^2 \\
\qquad + 560\mu_{3:x}^2\mu_{2:x} - 630\mu_{2:x}^4\}
\end{cases}
$$

In terms of the standard moments of the distributions these equations become

$$
(20)\begin{cases}
\alpha_{3:z} = \dfrac{1}{r^{1/2}}\alpha_{3:x} \\[2mm]
\alpha_{4:z} - 3 = \dfrac{1}{r}\left\{\alpha_{4:x} - 3\right\} \\[2mm]
\alpha_{5:z} - 10\alpha_{3:z} = \dfrac{1}{r^{3/2}}\left\{\alpha_{5:x} - 10\alpha_{3:x}\right\} \\[2mm]
\alpha_{6:z} - 15\alpha_{4:z} - 10\alpha_{3:z}^2 + 30 = \dfrac{1}{r^2}\left\{\alpha_{6:x} - 15\alpha_{4:x} - 10\alpha_{3:x}^2 + 30\right\} \\[2mm]
\alpha_{7:z} - 21\alpha_{5:z} - 35\alpha_{4:z}\alpha_{3:z} + 210\alpha_{3:z} = \dfrac{1}{r^{5/2}}\left\{\alpha_{7:x} - 21\alpha_{5:x}\right. \\
\qquad \left. - 35\alpha_{4:x}\alpha_{3:x} + 210\alpha_{3:x}\right\} \\[2mm]
\alpha_{8:z} - 28\alpha_{6:z} - 56\alpha_{5:z}\alpha_{3:z} - 35\alpha_{4:z}^2 + 420\alpha_{4:z} + 56\alpha_{3:z}^2 - 630 \\
= \dfrac{1}{r^3}\left\{\alpha_{8:x} - 28\alpha_{6:x} - 56\alpha_{5:x}\alpha_{3:x} - 35\alpha_{4:x} + 420\alpha_{4:x} + 56\alpha_{3:x}^2 - 630\right\}
\end{cases}
$$

If, without reference to subscripts, we write

$$
(21)\begin{cases}
\lambda_2 = \mu_2 \\
\lambda_3 = \mu_3 \\
\lambda_4 = \mu_4 - 3\mu_2^2 \\
\lambda_5 = \mu_5 - 10\mu_3\mu_2 \\
\lambda_6 = \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3 \\
\lambda_7 = \mu_7 - 21\mu_5\mu_2 - 35\mu_4\mu_3 + 210\mu_3\mu_2^2 \\
\lambda_8 = \mu_8 - 28\mu_6\mu_2 - 56\mu_5\mu_3 - 35\mu_4^2 + 420\mu_4\mu_2^2 \\
\qquad + 560\mu_3^2\mu_2 - 630\mu_2^4
\end{cases}
$$

the distribution of samples from an unlimited supply is defined, so far as moments through the eighth order are concerned, by the relations

$$
(22)\qquad\begin{cases}
M_z = rM_x \\[2mm]
\lambda_{n:z} = r\lambda_{n:x}
\end{cases}
$$

Working along a different line of approach, Thiele was the first to realize the importance of these λ functions. He made an extensive study of their unusual properties and was thus both directly and indirectly responsible for many important contributions to the theory of

mathematical statistics. These values of λ_i are the so-called "semi-Invariants of Thiele."

Again, we may write

$$
(23)\quad \begin{cases}
\gamma_3 = \alpha_3 \\
\gamma_4 = \alpha_4 - 3 \\
\gamma_5 = \alpha_5 - 10\,\alpha_3 \\
\gamma_6 = \alpha_6 - 15\,\alpha_4 - 10\,\alpha_3^2 \quad + 30 \\
\gamma_7 = \alpha_7 - 21\,\alpha_5 - 35\,\alpha_4\alpha_3 + 210\,\alpha_3 \\
\gamma_8 = \alpha_8 - 28\,\alpha_6 - 56\,\alpha_5\alpha_3 - 35\,\alpha_4^2 + 420\,\alpha_4 + 560\,\alpha_3^2 - 630
\end{cases}
$$

and observe that the shape of the distribution of samples is determined by the relation

$$
(24)\quad \gamma_{n:\bar{x}} = \frac{1}{r^{n/2-1}} \cdot \gamma_{n:x}
$$

which follows from equations (20).

The values γ_i are referred to as the 'standardized semi-invariants of Thiele."

If now r be permitted to approach infinity as a limit, we observe that in this limiting situation the *shape* of the distribution of samples is entirely independent of the shape of the parent population, since

$$
\lim_{r\to\infty} \gamma_{n:\bar{x}} = 0
$$

that is

$$
\begin{aligned}
\alpha_{3:\bar{x}} &= 0 \\
\alpha_{4:\bar{x}} - 3 &= 0 \\
\alpha_{5:\bar{x}} - 10\,\alpha_{3:\bar{x}} &= 0 \\
\alpha_{6:\bar{x}} - 15\,\alpha_{4:\bar{x}} - 10\,\alpha_{3:\bar{x}}^2 + 30 &= 0
\end{aligned}
$$

etc.

Thus the limiting distribution, which is called "the Normal Curve," must have the following properties:

$$
(25) \qquad \begin{cases} \alpha_{3:x} = 0 \\ \alpha_{4:x} = 1\cdot 3 \\ \alpha_{5:x} = 0 \\ \alpha_{6:x} = 1\cdot 3\cdot 5 \\ \alpha_{7:x} = 0 \\ \alpha_{8:x} = 1\cdot 3\cdot 5\cdot 7 \end{cases}
$$

The Theorem of Bernoulli

If p denotes the probability that an event will happen in a single trial and $q = 1 - p$ the probability that it will not happen in that trial, then the probability that the event will happen exactly x times during r trials is, by Bernoulli's Theorem

$$
(26) \qquad B_{r:x} = \binom{r}{x} q^{r-x} p^{x}
$$

From our point of view we need only regard the problem as one of sampling in which we withdraw samples of r variates from an infinite parent population, in which ,as per Table V, p designates the proportion of the variates which are zero in magnitude—the remaining variates being of unit magnitude. Then since

$$
\mu_{n:x} = pq\left\{ q^{n-1} + (-1)^n p^{n-1} \right\}
$$

we see from formulae (18) that

$$
(27) \qquad \begin{cases} M_x = rp \\ \mu_{2:x} = rpq \\ \mu_{3:x} = rpq\{q^2 - p^2\} \\ \mu_{4:x} = rpq\{q^3 + p^3\} + 3r^{(2)}p^2q^2 \\ \mu_{5:x} = rpq\{q^4 - p^4\} + 10\, r^{(2)}p^2q^2\{q^2 - p^2\} \end{cases}
$$

etc.

Poisson's Exponential Binomial Limit

If the probability that each of 1000 individuals die in one year were .5, then the expected number of deaths in such a group for one year would be 500. On the other hand, if the probability that each of 10,000 die in the year were .05 then the expected number of deaths would also be 500. Again $r=100000$ and $p=.005$ or $r=1000000$ and $p=.0005$ would give the same value. If we continue after this fashion to let r approach infinity and p zero, but in such a manner that the product $rp=M$ remains constant, then it can be shown quite readily that (26) becomes

$$\lim_{\substack{r\to\infty\\ p\to 0\\ rp=M}} B_{r:x} = \frac{e^{-M}M^{x}}{x!} \tag{28}$$

This is known as Poisson's Exponential Binomial Limit. For a Poisson distribution it follows from (27) that

$$\left\{\begin{aligned}
\mu_{2:x} &= M_x\\
\mu_{3:x} &= M_x\\
\mu_{4:x} &= M_x + 3M_x^2\\
\mu_{5:x} &= M_x + 10M_x^2\\
\mu_{6:x} &= M_x + 25M_x^2 + 15M_x^3\\
\mu_{7:x} &= M_x + 56M_x^2 + 105M_x^3\\
\mu_{8:x} &= M_x + 119M_x^2 + 409M_x^3 + 105M_x^4
\end{aligned}\right. \tag{29}$$

Substituting these values back in the definitions of the semi-invariants (formulae 21), we observe that for a Poisson distribution

$$\lambda_{n:x} = M_x \qquad (x = 2, 3, \cdots\cdots, 8) \tag{30}$$

Discussion of Results

So far as I know, no general method has been worked out which will permit one to express complex summations, such as those on pages

103, 104, in terms of moments. Moreover, I am unable at present to justify the use of the "sampling polynomials" for the moments of the samples of an order higher than the eighth. Laborious computations have established the fact that the apparent law of the sampling polynomials holds for the first eight moments, and hence we have a simple method at our disposal of writing down expressions for these moments of samples withdrawn from finite parent populations. A study of these sampling polynomials should reveal an entirely different approach to the problem. This is but one of many interesting problems of mathematical statistics that require further investigation.

Although we utilized the results of sampling from a limited supply to obtain corresponding formulae for sampling from an unlimited supply, nevertheless it can be shown that for $s=\infty$ a simple method exists for expressing the moments in terms of the moments, $\mu_{n:x}$, as in formulae (18). Moreover, this law holds for any positive integer, n.

Thus

$$\mu_{20:s}=\frac{20!}{20!}r^{(1)}\mu_{20:x}+\frac{20!}{18!\,2!}\,r^{(2)}\mu_{18:x}\,\mu_{2:x}+\cdots\cdots$$

$$+\cdots\cdots\frac{20!}{9!\,7!\,4!}\,r^{(3)}\mu_{9:x}\,\mu_{7:x}\,\mu_{4:x}+\cdots\cdots$$

$$+\cdots\cdots\frac{20!}{(4!)^2(3!)^4}\,r^{(6)}\,\frac{\mu^2_{4:x}\,\mu^4_{3:x}}{2!\,4!}+\cdots\cdots$$

Since formulae, such as (3a) and (4a) are based on multinomial considerations, the rule for writing down the values of $\mu_{n:s}$ is valid for any value of n, when $s=\infty$.

Proceeding after this fashion, one can show that corresponding to formulae (25) one can write for the limiting distribution, referred to as the Normal Curve,

$$(31)\qquad \begin{cases} \alpha_{2n+1:s}=0 \\[6pt] \alpha_{2n:s}=\dfrac{(2n)!}{2^n(n!)} \end{cases}$$

And since the function

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \tag{32}$$

satisfies the above conditions, we say that (32) is the equation of the Normal Curve. In the Theory of Least Squares this equation is usually* developed on the so-called Hagen's hypothesis, that is "An error is the algebraic sum of an indefinitely great number of small elementary errors which are all equal, and each of which is equally likely to be positive or negative."

From the results that we have obtained it appears that it is not necessary to impose the restrictions that the elementary errors are all equal and that positive and negative values are equally likely. It is necessary only that

(1) the number of elementary errors be infinite, although of an order less than that of the number of errors in the parent population.

(2) the errors be independent. This restriction is really involved in our assumption that in evaluating summations, each of the s variates of the parent population occurs exactly as many times as every other variate.

Otherwise, the limiting shape of the distribution of samples is independent of the shape of the parent distribution. The fact that tables II and IV, arising from parent distributions that are so extremely abnormal, exhibit distributions of samples that are fairly normal, seems to bear out our point in spite of the fact that we employed in each instance a small value of r, i. e. twenty-five.

*See Merriman's Method of Least Squares. John Wiley and Sons, New York City.